THE UNIVERSITY *of* EDINBURGH

Chromosome folding and organisation across different organisms: a Molecular Dynamics study

Maria Carolina Figueirinhas Pereira

Doctor of Philosophy
The University of Edinburgh
March 2019

Abstract

Polymer models have long been used to study the properties and behaviour of DNA, however the principles behind chromosome folding and organisation remain elusive. In this thesis we will analyse the contributions of different mechanisms driving genome compaction, such as macromolecular crowding, and interactions with different DNA-binding proteins.

For this we will use Molecular Dynamics simulations of coarse-grained polymer models of both bacterial and eukaryotic DNA, together with methods of equilibrium and non-equilibrium Statistical Mechanics. Our study is motivated by recent experiments probing the compressional elasticity and dynamics of single bacterial chromosomes confined in a cylindrical pore, and by new high-throughput experimental techniques that capture the genome conformation in living cells.

We start by looking at the properties of bacterial DNA. Our major contribution is the quantification of the effect of different compaction mechanisms on the DNA response to compression. We conclude that crowding proteins in particular strongly affect both the compression curves and the expansion dynamics. We also give evidence of a novel popping-off kinetic regime during expansion, where DNA-binding-proteins detach one by one leading to a slow unfolding dynamics. These results are robust with respect to changes in protein size and particle charge.

We then turn to the study of the 3-D spatial organisation of human chromosomes. We conciliate two previously competing viewpoints regarding the mechanisms driving chromosome conformation in human cells. Thus, we show that transcription factors organise active and repressed chromatin, and establish long-range interactions leading to active/inactive domain phase separation, whereas chromatin loop extruding proteins (cohesin) are necessary to form domains within inert chromatin, which lacks binding sites for molecular bridges. We also show that a version of the model where chromatin loop extruders move diffusively rather than actively works equally well - this is important in view of single molecule experiments which have yet to find a motor activity in cohesin on chromatin fibres. Our model predicts the chromosome structure captured in experiments and the effect of various protein knock-outs.

We finish by studying the nuclear organisation of fruit fly chromosomes and the subsequent formation of nuclear bodies. By modelling all chromosomes in the nucleus of an haploid cell, we are able to predict the structural and dynamical properties of the whole genome. We show that the formation of nuclear bodies is linked to genome reorganisation after mitosis (cell division). Our model predicts the dynamics and size distributions probed experimentally of such nuclear bodies. Importantly, we show that the large-scale chromosomal organisation is tightly dependent on the chromosomal conformation just after mitosis.

Lay Summary

For over 50 years physicists have worked on theoretical models in order to study and characterise the behaviour of polymers. One of the most studied polymers since its discovery is DNA – the molecule that contains the genetic information necessary to run all biological functions in living organisms. However the mechanisms driving the compaction and organisation of a DNA molecule, that in human cells is about two meters long, so that this fits inside a micrometer-sized cell nucleus, still remain somewhat elusive.

The aim of this Thesis is then to perform a comprehensive analysis of the contributions of different mechanisms responsible for DNA folding and compaction. For this we will use computer simulations to model the behaviour of DNA, in the presence of DNA-binding and non-binding proteins, across different organisms.

We start by studying the elastic and dynamic responses of bacterial DNA to compression, and how these are affected in the presence of proteins. We do so by modelling a single DNA molecule inside a nano-channel that can be compressed using a piston. We conclude by giving evidence of a novel "popping-off" protein behaviour that emerges during DNA expansion, after initial compression, where DNA-binding-proteins detach one by one leading to a slow unfolding dynamics.

We then turn to the study of the chromosome folding properties in human cells. We conciliate two previously competing viewpoints regarding the key mechanisms responsible for the spatial structure of chromosomes. Thus, we show that proteins that act like molecular bridges between different DNA segments are responsible for organising gene-rich chromosome regions, whereas proteins that act like molecular handcuffs, which bring DNA segments together and then slide creating chromosome loops, are responsible for organising gene-poor chromosome regions. The simulation results are in good agreement with experimental data for the 3D chromosome structures, and the model can even predict the results from protein knock-out experiments.

We finish by studying the nuclear organisation of chromosomes and proteins in fruit fly cells. We do so by modelling all chromosomes and proteins inside a

spherical nucleus. We show that the formation of protein clusters is linked to the chromosomes' nuclear organisation after cell division, and we characterise the properties of such clusters. The model predictions are in good agreement both with experiments probing the 3D chromosome structures and with experiments probing the formation of nuclear protein clusters.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Parts of this work have been published in:

- M. C. F. Pereira, C. A. Brackley, J. S.Lintuvuori, D. Marenduzzo, E. Orlandini, *Entropic elasticity and dynamics of the bacterial chromosome: a simulation study*, J. Chem. Phys., **147**, 044908 (2017)

- M. C. F. Pereira, C. A. Brackley, D. Michieletto, C. Annunziatella, S. Bianco, A. M. Chiariello, M. Nicodemi, D. Marenduzzo, *Complementary chromosome folding by transcription factors and cohesin*, *sub judice* (bioRxiv 305359)

I gratefully acknowledge the help of Dr. C. A. Brackley and Dr. D. Michieletto in retrieving experimental data such as ChIP-seq and HMM profiles, and HiC interaction maps.

(*Maria Carolina Figueirinhas Pereira, March 2019*)

Maria Carolina Figueirinhas Pereira, March 2019

Acknowledgements

I would like to start by thanking my beloved husband for his unconditional love and support during this past four years, for being my guiding light in the most difficult moments, for giving me the greatest joys in life, and most importantly for constantly helping me become a better person!

I would like to thank my family – my parents, my sisters, and my in-laws – for all their support, and for always welcoming us home with a true happiness and warm love.

I would like to thank all our old friends from home for their faithful friendship, and our new friends we made in Scotland with whom we learned so much during this journey. I deeply thank them for filling our hearts with wonderful memories.

I would also like to express my sincere gratitude to my supervisor Davide, who helped and taught me so much during my PhD studies. It was a real pleasure to work with someone with a such a great knowledge and who is always extremely kind and willing to help. I also owe my sincere thanks to the other members of our group, specially to Chris and Davide Michieletto who always promptly helped me whenever I needed.

Finally, and most importantly, I would like to thank God with all my heart for making this journey possible, and for being my refuge, my hope, and my life!

Contents

List of Figures

List of Tables

Introduction 1

The building block of all living organisms is the cell – it is the smallest unit of life that can replicate independently and where all key biological processes take place [1]. The necessary information for running the cellular activities is encoded in the famous macromolecule named DNA (Deoxyribonucleic acid). A fully stretched DNA molecule, isolated in its thread-like form, can be three (bacterial DNA) to five (human DNA) orders of magnitude longer than the size of the cell where it is embedded in. That is why Nature promotes folding and organisation of the genetic material into more compact structures – the chromosomes [22].

However, DNA cannot be compacted too tightly. The information stored along the DNA's length has to remain accessible for transcription, binding of regulatory proteins, replication, and DNA damage repair [22]. Therefore, the cellular activity is strongly linked with the structure and dynamics of chromosomes [23, 142], and studying the mechanisms of compaction and organisation of the genetic material becomes imperative for the comprehension of the cellular processes, either for the purpose of developing new biotechnological techniques, such as plant and animal fertility methods, or new disease treatments.

The DNA molecule was first isolated by Miescher in 1869, but its molecular structure was only discovered $\sim$ 80 years later, in 1953, by Crick, Watson and Franklin. The basic form of the DNA molecule is illustrated in Figure 1.1. It consists of two strands coiled around each other forming a right-handed double helix, each strand being composed of a series of units called nucleotides. The double helix form of the DNA is stabilised by the pairing of the nucleotides of the two strands, with each complementary pair of nucleotides being referred to as a base pair (bp). DNA runs the activities of cells by specifying the composition and structure of proteins, which are responsible for the regulation of all cellular processes. This information is encoded in specific base-pair sequences called genes. The other DNA sequences that do not code for proteins are

known as noncoding or "junk" DNA sequences, but contrary to what the name suggests these sequences also have important functions, like in gene expression regulation [1]. By influencing the overall DNA spatial organisation, noncoding DNA helps determining which genes remain active or inactive. For instance, genes embedded in DNA regions that are highly compacted cannot be accessed by the cell machinery in order to be transcribed and, thus, are effectively inactive. On the other hand, genes located in more unfolded regions are easily transcribed, thus being in an active state.

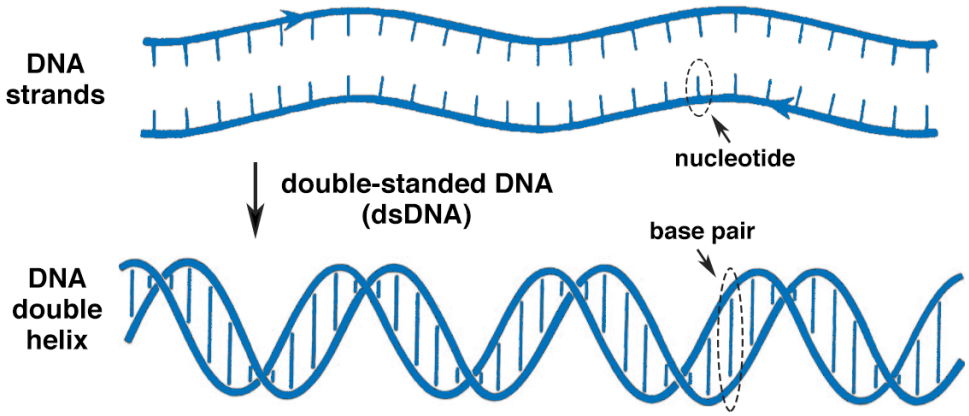

Figure 1.1 *Schematic representation of a DNA segment: (top) denatured DNA strands; (bottom) the DNA double helix. Adapted from [22].*

As mentioned above, due to the small cellular dimensions, DNA is folded into chromosomes. Depending on how chromosomes organise inside the cell, one can have prokaryotic or eukaryotic cells. In prokaryotic cells (Fig. 1.2) the DNA molecule describes a closed loop and is dispersed in the cellular medium. However it does not occupy the entire cell volume, being instead somewhat constrained to a smaller volume known as the nucleoid. In eukaryotic cells (Fig. 1.3) DNA molecules are chain-like polymers and are confined in a nucleus, being well separated by a membrane from the cellular environment. Genome-related activities, such as DNA transcription, replication and repair, all happen within the nucleus.

In spite of decades of study of the cellular components and their function, the mechanisms that lead to the balanced compaction and organisation of DNA are not yet fully understood. For prokaryotes, like bacteria, three main mechanisms have been proposed: the effect of entropic forces due to macromolecular crowding [58, 60, 85], the binding of nucleoid-associated proteins [17, 36],

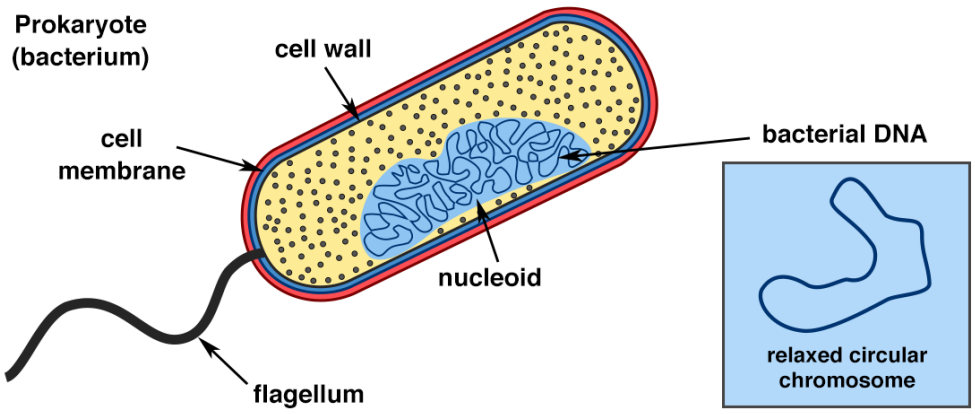

Figure 1.2 *Diagram showing the basic structure of a bacterial cell. The bacterial DNA is condensed in the nucleoid, without being confined by an additional membrane.*

and the partition of the DNA molecule into topologically isolated supercoiled domains [22, 129]. However their comprehensive study has been hindered by the experimental difficulties when dealing with prokaryotic cells, due to their small size but complex molecular and biochemical composition. For eukaryotes, the mechanisms organising DNA are well understood at the smallest length scales (see Fig. 1.3): DNA first wraps around protein complexes, composed of eight histone proteins (histone octamers), forming nucleosomes; this "beads on a string" DNA structure yields the so-called 10 nm chromatin fibre; at physiological concentrations ($\sim$ 150 mM of a monovalent salt), the nucleosomes assume a stacked conformation so that DNA becomes even more compacted into a 30 nm chromatin fibre [1]. However, there is not yet a comprehensive picture of the mechanisms driving large-scale chromatin organisation, through the different stages of the cell life cycle. This is again due to experimental difficulties in singling out the key organising cell machinery, since this requires extensive high resolution observations.

Therefore, the aim of the work presented in this Thesis is to shed new light on the effect of different mechanisms, such as macromolecular crowding and interactions with different types of DNA-binding proteins, on the compaction/folding properties of DNA across various organisms. This Thesis is largely inspired by recent experimental advances that allow for the manipulation of single bacterial chromosomes and direct measurement of their compressional elasticity and dynamics [104], and by new experimental high-throughput techniques that yield genome-wide maps of chromosomal interactions in living cells, thus allowing for

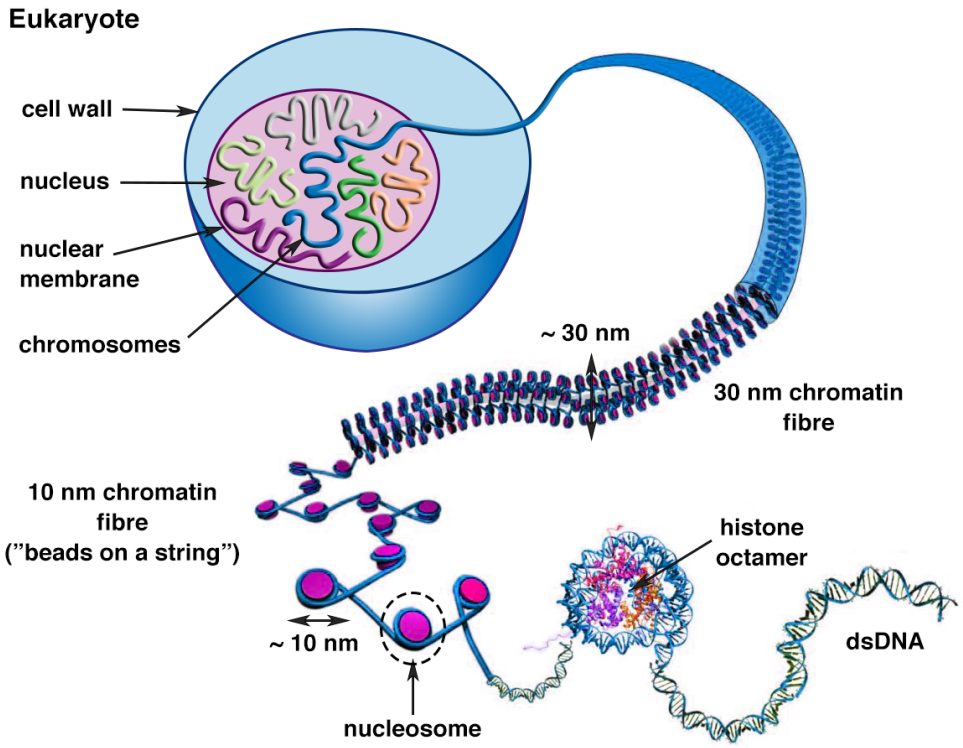

Figure 1.3 *Diagram showing the basic structure of a eukaryotic cell. In eukaryotes, DNA is confined in the nucleus by the nuclear membrane. Specialised proteins help packaging/folding the DNA, thus providing increasingly higher levels of genome organisation. Adapted from [53].*

the reconstruction of the spatial organisation of chromosomes in the nucleus [136].

The approach used in this work is the one of polymer physics. Specifically, DNA will be modelled as a polymer at a coarse-graining level that allows for a simple representation of DNA, while retaining its key physical elements. This approach not only allows us to make use of well-known theoretical results to predict the behaviour of DNA, either in solution or in confinement, but also provides us with the flexibility to use the same simple model to study various systems at different length scales, simply by adjusting a few parameters.

There are two different ways of designing the suitable polymer model: top-bottom, and bottom-up. According to the former, the model is first fitted to existing experimental data, for instance probing the 3D chromosome structure, so as to find the optimised parameters and force fields for the polymer interactions which can recreate these data. According to the latter, the key systems' features to be taken into account are first chosen based on basic biological assumptions,

such as the polymer flexibility and the chromatin state sequence along the chromosome, and then the model is designed to include these. The work in this thesis follows the second method (bottom-up). This way, the obtained results emerge from a fitting-free model, thus yielding predictive conclusions.

This Thesis will be divided in three parts, where different versions of the same polymer model will be used to study the properties of genome folding in three different biological systems. The first part will be devoted to the study of the elastic and dynamic properties of bacterial chromosomes, and how these are affected by explicit proteins. This will be the simplest system considered, consisting of a short DNA polymer that only interacts with proteins non-specifically.

The second part will focus on the human genome with the purpose of investigating what mechanisms drive large-scale chromatin folding during interphase. Interphase is the stage of the cell life cycle in which cells spend the majority of their time. It is the phase when cells actually "live", i.e., when they obtain nutrients, perform all cellular processes, and grow. During this stage, chromosomes adopt a less compact structure, promoting the transcription of specific genes. It is, therefore, during interphase that chromatin conformation and gene expression are intimately linked, hence the focus of this part of the Thesis on this cell stage. The main goal will be to conciliate two competing viewpoints regarding different chromosome folding mechanisms, by thoroughly studying their effect on the folding of a long chromosome segment with a varied genomic landscape, i.e., that contains gene-rich and gene-poor regions.

The third part of this Thesis will turn the focus from studies with only one chromosome to a case study with multiple chromosomes. The chosen system will be the *Drosophila melanogaster* (fruit fly) genome. The fruit fly constitutes a good model system, from the simulations point of view, since it has much less and shorter chromosomes than human cells. The aim here will be to study the nuclear organisation of chromosomes and proteins after cell mitosis. Cell mitosis is the stage in which the cell undergoes division into two daughter cells. During this stage, there are two copies of each chromosome and all chromosome pairs further condense, assuming an X- or V-shape. These structures are then divided in half so that a copy of each chromosome goes to either side of the cell. Once the genetic material is correctly partitioned, the cell divides at the centre giving origin to two identical new cells. Therefore, right after mitosis, the chromosomes

in each daughter cell are still highly condensed, starting to adopt a less structured conformation as the cell enters in interphase. This part of the Thesis will then be devoted to understanding how chromosomes' structure reorganises after cell division, and how the nuclear arrangement of proteins affects this process (and is in turn affected by it), by modelling all chromosomes of an haploid cell[1] in a spherical nucleus.

Finally, the work presented in this Thesis will be structured as follows:

In Chapter 2 I will give an overview of the polymer scaling laws for linear polymers, in solution and under confinement. I will also review the qualitative behaviour of polymers when subjected to crowded environments due to macromolecules or proteins, and when interacting with polymer-binding proteins.

In Chapter 3 I will describe the polymer model and simulation details that serve as a basis (are common) to the different systems that will be studied in the first, second and third part of the Thesis.

In Chapter 4 I will study the physical properties of bacterial chromosomes by considering a simulation system closely related to a recent experimental study scheme [104]. In that study, single bacterial chromosomes were trapped in nanopores, and then compressed using a micropiston. This allowed to measure the entropic force exerted by the DNA on the piston as a function of the DNA extension along the pore, giving a direct estimate of the DNA entropic spring constant. Following the idea of the experimental set-up, I will consider a chromosome trapped in a cylindrical pore, and will start by quantifying the effect of the presence of crowding proteins and DNA-binding proteins on the entropic response of the DNA. I will then investigate the dynamical behaviour of DNA while expanding along the pore after having been compressed, and once again address the effect of proteins. I will finish by describing an interesting protein "popping-off" dynamics of DNA-binding proteins that emerges for specific DNA-protein affinity strengths.

In Chapter 5 I will address the role of two types of proteins – transcription factors and cohesin – on the spatial organisation of human chromosomes in interphase. These two protein types are the basis of two competing models

[1] In common diploid cells there are two chromosome sets, inherited one from each parent. An haploid cell is a cell containing only half of the usual number of chromosomes.

for genome organisation. I will compare their effect on a chromosome segment containing both gene-rich and gene-poor regions, by starting to implement the basic forms of each model in separate simulations, and ending by looking at a combination of both models. I will then explore a more complex version of the model by considering a mechanism for protein switching between DNA-binding and non-binding modes, which drives the whole system away from equilibrium. This switching mechanism is inspired by recent experiments that suggest more dynamical protein clusters [79, 134]. I will then investigate an alternative model for cohesin based on a "diffusive" rather than active mechanism, since to date evidence for a motor activity has not been found. And finally I will test the proposed models by predicting the effect of two protein knock-outs.

In Chapter 6 I will investigate the spatial reorganisation of chromosomes and proteins, after cell division, in the nucleus of *Drosophila melanogaster* cells. I will start by looking at how chromosomes relax after mitosis, analysing the effect of the chromosomes' initial conformation on the long term nuclear organisation. I will then describe the formation of nuclear bodies, particularly those composed of Polycomb Group (PcG) proteins. This is inspired by recent super-resolution microscopy experiments [138], that allowed for the imaging and characterisation of PcG bodies in the nucleus of *Drosophila* cells. I will proceed with an analysis of the nuclear distribution, predicted by the proposed model, of active, inactive and silenced chromatin regions, as well as of the different modelled nuclear bodies. And finally I will address the effect of chromatin-nuclear wall interactions.

2 DNA from a polymer physics' perspective

For over 60 years, physicists have worked on theoretical models [34, 38, 42, 118] to characterise the behaviour of polymers. These are essentially long molecules that play a key role in diverse systems, from rubbers to living cells. Due to their flexible and chain-like nature, their behaviour cannot be studied using conventional mechanistic models. Instead, statistical approaches are used to infer the macroscopic properties of polymer systems, by deriving universal scaling laws.

These scaling relations create a robust framework, such that they are used to study apparently very distinct systems, like chewing gum and DNA. Therefore, in this Thesis I will make use of such powerful approaches to explore the macroscopic properties of DNA. For that, I will overlook the atomistic details of DNA, and describe a DNA molecule using a coarse-grained polymer model, where DNA is viewed as a linear polymer composed of chained monomers.

In what follows, I will review the main theoretical approaches developed to infer the elastic and dynamic behaviour of linear polymers. I will start by reviewing the scaling laws for polymers in solution and under confinement, and end by describing the effect of protein crowding and polymer-binding proteins on the dynamic and conformational polymer properties.

2.1 DNA in solution

2.1.1 The ideal polymer

The simplest model used to infer the statistical properties of a polymer is the *freely jointed chain* (FJC), which describes the polymer simply as a chain of non-

interacting monomers with a fixed monomer-monomer distance (see Fig. 2.1(a)). One can define the bond variable between two consecutive monomers as

$$\boldsymbol{\tau}_i = \boldsymbol{r}_i - \boldsymbol{r}_{i-1}, \tag{2.1}$$

where $\boldsymbol{r}_i$ is the position of monomer i and $|\boldsymbol{\tau}_i| = \sigma$ the monomer-monomer distance, also known as the Kuhn length l_K. In the FJC picture monomers are positioned according to an ideal random walk, so that they are allowed to overlap each other. This means that the bond variables are independent and can take any orientation irrespective of the other bonds. This absence of correlation between different bonds translates to the following condition:

$$\langle \boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_j \rangle = \sigma^2 \delta_{ij}, \tag{2.2}$$

where $\langle \dots \rangle$ stands for the average over the ensemble of all possible polymer chain configurations. Eq. (2.2) leads to the well known scaling relation for the end-to-end distance of an N-monomer-long freely jointed chain, $\boldsymbol{R}$:

$$\langle \boldsymbol{R}^2 \rangle = \left\langle \left(\sum_i^N \boldsymbol{\tau}_i \right)^2 \right\rangle = \sum_{i,j}^N \langle \boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_j \rangle = \sigma^2 N \quad \Rightarrow \quad R \approx \sigma N^{1/2}. \tag{2.3}$$

Thus the end-to-end distance depends on the number of monomers as

$$R \approx \sigma N^{\nu}, \quad \text{with} \quad \nu = 1/2. \tag{2.4}$$

A related quantity is the free energy of an ideal chain with the two ends held at a distance R. Since there are no interactions between monomers, the free energy is set by the chain entropy. According to Boltzmann, the entropy of the chain is simply given by $S(R) = k_B \ln \Omega(R)$, where $\Omega(R)$ is the number of possible configurations for a chain with an end-to-end distance R. Therefore $\Omega(R)$ is proportional to the probability of finding a chain with the ends held at R, which is a Gaussian: $P(N, R) \approx \exp(-R^2/\sigma^2 N)$ [118]. Therefore, $S(R) \approx -k_B R^2/\sigma^2 N$, and the free energy becomes:

$$\mathcal{F}(R) = -TS(R) \approx k_B T \frac{R^2}{\sigma^2 N}. \tag{2.5}$$

From here it follows that the force required to perturb an ideal chain is given

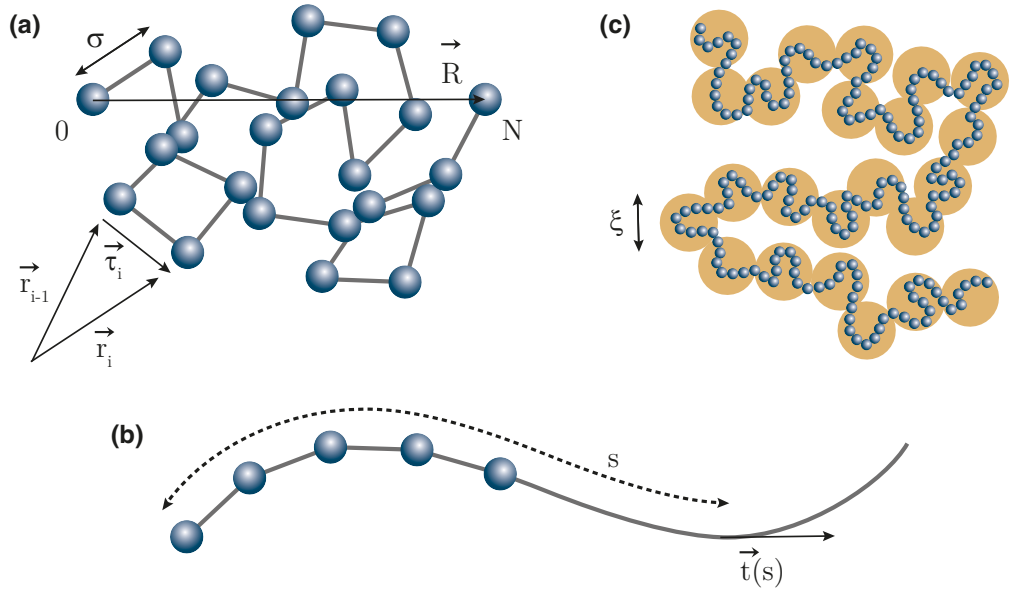

Figure 2.1 *Different representations of a polymer. (a) Freely Jointed Chain (FJC): the polymer is described as a chain of monomers connected by rigid bonds that freely rotate. Monomers and bonds can cross without penalty. The polymer conformation is described by an ideal random walk. (b) Worm-Like Chain (WLC): the polymer is described as a smooth curve. The monomer position is given by the length s along the polymer and the bond orientation is given by the tangent vector. (c) Polymer blob approach: coarse-grained description where the polymer is described as a chain of blobs, each containing g monomers.*

by

$$f = \frac{\partial \mathcal{F}(R)}{\partial R} \approx k_B T \frac{R}{R_0^2}, \tag{2.6}$$

where R_0 is the end-to-end distance of the unperturbed ideal chain as in Eq. (2.4). The force is linear in the polymer extension R, meaning that the ideal chain behaves as a Hookean spring, and Eq. (2.5) is in fact an elastic free energy. Since the behaviour of the chain is governed by its entropy, the ideal chain is also said to behave like an entropic spring. The result in Eq. (2.6) holds only for deformations that do not significantly alter the chain statistics, i.e., that do not break the Gaussian approximation.

2.1.2 Worm-like polymers

A key property of a polymer like DNA is its resistance to bending [1, 22]. In fact, DNA is a semi-flexible polymer, and bending the double helix incurs an

energetic penalty. A refined model of the ideal chain that takes such property into account is the worm-like chain. Suggested by Kratky and Porod in the late 40s, this model has been used in a variety of studies on the behaviour of DNA subjected to stretching forces [77, 84, 87, 116], giving good predictions of the experimental curves [12, 137].

According to this model, successive bonds between monomers all point roughly in the same direction, and the polymer conformation can be described as a smooth curve (see Fig. 2.1(b)). The bending rigidity of the chain is then quantified by a persistence length l_p. This quantity essentially gives a measure of the length scale along the polymer backbone, for a given temperature T, beyond which correlations in the local orientation decay:

$$\langle \boldsymbol{t}(s) \cdot \boldsymbol{t}(s') \rangle = e^{-|s-s'|/l_p}, \tag{2.7}$$

where $t(s)$ is the tangent to the polymer contour at chain length s. For hydrated double-stranded B-DNA *in vivo* (where the physiological concentration is ∼150 mM [3]) $l_p = 50$ nm. In this continuum formalism, the end-to-end distance of a chain with contour length $L_c = N\sigma \equiv N l_K$ is defined as

$$\boldsymbol{R} = \int_0^{L_c} \boldsymbol{t}(s) ds. \tag{2.8}$$

Following the reasoning made for the ideal chain, R can be determined as $\sqrt{\langle \boldsymbol{R}^2 \rangle}$:

$$\begin{aligned} \langle \boldsymbol{R}^2 \rangle &= \int_0^{L_c} ds \int_0^{L_c} \langle \boldsymbol{t}(s) \cdot \boldsymbol{t}(s') \rangle ds' = \int_0^{L_c} ds \int_0^{L_c} e^{-|s-s'|/l_p} ds' \\ &= 2 l_p L_c \left[1 + \frac{l_p}{L_c} \left(e^{-L_c/l_p} - 1 \right) \right]. \end{aligned} \tag{2.9}$$

For $L_c \gg l_p$, $\langle \boldsymbol{R}^2 \rangle \approx 2 l_p L_c = 2 l_p l_K N$, and one recovers the ideal chain scaling:

$$R \approx l_K N^{\nu}, \quad \text{with} \quad \nu = 1/2, \tag{2.10}$$

where $l_K = 2 l_p$ is the effective ideal random walk monomer-monomer distance (Kuhn length).

In summary, for polymer segments shorter than l_p the chain behaves like a stiff beam, whereas for segments longer than l_p thermal fluctuations alone can induce

bending of the chain. For length scales much beyond the persistence length, the polymer actually behaves like a freely jointed chain and its behaviour can be described by the scaling relations revised in the previous section.

Real polymers, like DNA, however are far from behaving like ideal chains. The freely jointed chain model can be easily adapted to describe more realistic polymers, where monomers do in fact have a finite size and hence cannot occupy the same volume. The inclusion of excluded volume interactions leads to the so-called *self-avoiding polymer* model.

2.1.3 More realistic polymers

One of the most successful approaches to describe the behaviour of self-avoiding polymers was proposed by Flory [42], and consists of minimising the Helmholtz free energy of a chain with the two ends held at $R = |\boldsymbol{R}|$:

$$\mathcal{F}(R) = U(R) - TS(R), \tag{2.11}$$

where $U(R)$ is the chain internal energy associated with the inter-monomer interactions, and $S(R)$ the entropy of the chain.

The internal energy can be estimated as the repulsion energy between all pairs of monomers that happen to occupy the same volume. Given the monomer concentration

$$c = \frac{N}{R^d}, \tag{2.12}$$

where R^d is the volume in d-dimensions where monomers are to be found, the probability of finding a monomer in a small volume v is $\sim vc = vN/R^d$, and of finding two monomers in the same volume v is $\sim (vN/R^d)^2$. The total number of interacting pairs is given by summing $(vN/R^d)^2$ for all the R^d/v volumes v within the volume R^d: vN^2/R^d. Considering $v \sim \sigma^d$, the overall repulsion energy is $U(R) \approx k_B T\ \sigma^d N^2/R^d$, with k_B being the Boltzmann constant.

The entropy is simply given by the ideal chain entropy (see Sec. 2.1.1), so that the free energy becomes

$$\frac{\mathcal{F}(R)}{k_B T} \approx \frac{\sigma^d N^2}{R^d} + \frac{R^2}{\sigma^2 N}. \tag{2.13}$$

Minimising $\mathcal{F}(R)$ with respect to R leads to the self-avoiding polymer's scaling

in d-dimensions:

$$R \approx \sigma N^{3/(d+2)} \quad \Rightarrow \quad R \approx \sigma N^{\nu}, \quad \nu = \frac{3}{d+2}, \tag{2.14}$$

which agrees with the known results $\nu = 1$ for $d = 1$ and $\nu = 3/4$ for $d = 2$, and is also very close to the best estimate for $d = 3$ that gives $\nu \approx 0.588$ [76]. For $d = 4$ one recovers the value for the FJC $\nu = 1/2$. However for $d > 4$ this result predicts an exponent $\nu < 1/2$, which is unrealistic since exclusion volume interactions cannot cause the polymer to become more compact than the ideal chain. Therefore, for $d > 4$ one should expect the FJC exponent $\nu = 1/2$ so that the general size exponent becomes:

$$\nu = \begin{cases} \frac{3}{d+2} & \text{for } d \leq 4, \\ 1/2 & \text{for } d > 4. \end{cases} \tag{2.15}$$

This result is also called the Flory exponent ν_F, and $R_F \approx bN^{\nu_F}$ the Flory radius.

It is surprising that in most practical applications Flory's formula can be considered to be exact. In fact, Eq. (2.15) is the result of a remarkable cancellation of errors, since both repulsive and elastic energies are overestimated. Works that focused on deriving a more accurate expression for only one of those terms, actually turned out to give a worse final result for ν.

2.1.4 The blob scaling picture

In the late '70s [34], de Gennes went one step further and developed another approach that relies on the notion of polymer blobs and uses Flory's theory as a basis. The so-called "blob scaling approach" creates a simple framework to study the effects of internal interactions and external perturbations on the polymer statistical properties.

As the name itself suggests, this theory visualises a polymer chain as a succession of units, or "blobs", of size ξ (see Fig. 2.1(c)). Inside each blob, the chain does not interact with other chains, neither is it affected by external perturbations, like confinement or applied forces. Therefore, in each blob the chain properties must be governed by excluded volume interactions, meaning

that one can expect the scaling behaviour described in Section 2.1.3:

$$\xi \approx \sigma g^{\nu}, \tag{2.16}$$

where g is the number of monomers inside each blob.

The blob scaling picture becomes specially useful to infer the scaling behaviour of polymers subjected to external forces or confinement.

2.2 DNA under confinement

As mentioned in Chapter 1, DNA is highly confined in living cells. Moreover, advances in technologies for probing the elastic properties of single molecules rely on confining the DNA in cylindrical nanochannels with width $\sim$ 100nm [68, 104, 112, 114]. Therefore it is crucial to ask how confinement changes the scaling laws reviewed in the previous sections.

2.2.1 DNA trapped in a nanopore

A polymer confined in a cylindrical pore, of width $D < R_F$, tends to align along the cylinder axis (see Fig. 2.2). However, this tendency is opposed by the chain entropy, which favours coiled conformations. The polymer chain ends up "breaking" into a linear string of blobs of size $\xi = D$. At the blob scale, the effects of confinement are weak, and so the polymer behaves as an unconstrained self-avoiding chain obeying the scaling relation in Eq. (2.16): $\xi \equiv D \approx \sigma g^{\nu}$. The number of monomers per blob is then given by $g \approx (D/\sigma)^{1/\nu}$. But at larger length scales ($> \xi$), the blobs behave as hard spheres aligned in a one-dimensional array.

According to this picture, the relevant quantity to infer the chain size is the longitudinal chain extension L (see Fig. 2.2(a)). In the absence of further external forces, the equilibrium chain size L_0 is given by the number of blobs N/g:

$$L_0 = \frac{N}{g} D \approx N\sigma \left(\frac{\sigma}{D}\right)^{\frac{1}{\nu}-1}. \tag{2.17}$$

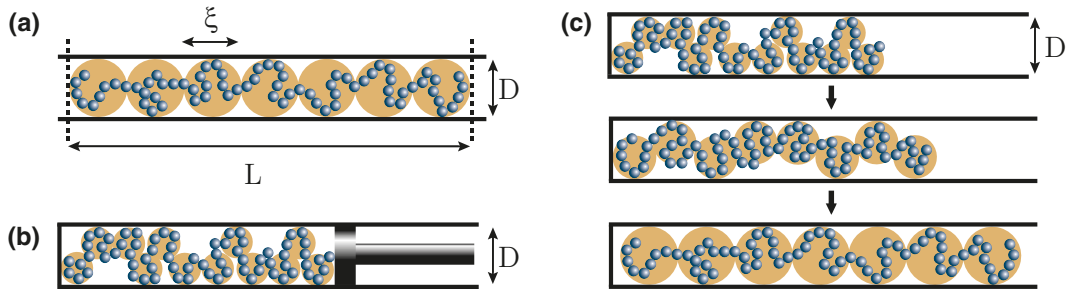

Figure 2.2 *(a) A polymer confined in a cylindrical pore with width $D < R_F$ "breaks" up into a linear string of blobs of size ξ. (b) Upon strong compression, the polymer blobs break into smaller ones. (c) If the polymer is then allowed to freely expand, the system goes through different regimes and the picture in (a) is eventually recovered.*

2.2.2 Response to applied forces

In Section 2.1.3, we have seen the Flory's free energy of an unconfined chain with an end-to-end distance R (Eq. (2.13)). However, in the early 2000s, Jun et al. showed that a "renormalised" free energy could be constructed for polymers under cylindrical confinement [59, 63]. This approach takes advantage of de Gennes' blob picture, and consists of considering the confining cylinder to be an effective one-dimensional space where the blobs can move. In other words, the polymer is effectively described as a one-dimensional chain of impenetrable blobs. According to this, Eq. (2.13) can be rewritten as a function of the chain extension L by simply setting $d = 1$, and rescaling $\sigma \to D$ and $N \to N/g$:

$$\frac{\mathcal{F}_{\text{cyl}}(L, D)}{k_B T} \approx \frac{D(N/g)^2}{L} + \frac{L^2}{D^2(N/g)}. \tag{2.18}$$

This renormalisation recovers the expected equilibrium chain size, for which $\mathcal{F}_{\text{cyl}}$ is minimised: $L_0 \approx (N/g)D \approx N\sigma(\sigma/D)^{1/\nu-1}$ (see Eq. (2.17), where ν is the Flory exponent given by Eq. 2.15). The free energy can then be written in the short form:

$$\mathcal{F}_{\text{cyl}}(L, D) \approx \frac{k_B T}{D}\left[\frac{L_0^2}{L} + \frac{L^2}{L_0}\right]. \tag{2.19}$$

Eq. (2.19) remains valid unless the chain is compressed to the point where $L/L_0 \ll 1$, or extended to the point where $L/L_0 \gg 1$, beyond which the chain blobs break into smaller ones (Fig. 2.2(b)) and the picture in Figure 2.2(a) no longer holds.

The important feature of this rescaled free energy is that it allows to derive a universal scaling relation between the chain extension L and the force exerted by the polymer when compressed/extended from its equilibrium configuration:

$$f = \frac{\partial \mathcal{F}_{\text{cyl}}(L, D)}{\partial L} \quad \Rightarrow \quad Df = k_{\text{chain}}\, k_B T \left[\left(\frac{L}{L_0} \right) - \left(\frac{L}{L_0} \right)^{-2} \right], \tag{2.20}$$

where k_{chain} is the dimensionless chain spring constant. Therefore, the force-extension relation does not depend on the number of monomers per blob, hence this formula is well suited to study the chain behaviour when g is not constant.

For strong compression ($L/L_0 \ll 1$), the polymer enters in the semidilute regime and can no longer be described by a linear chain of polymer blobs of size D. Instead, it is described as a chain of blobs of size $\xi \approx \sigma g^{\nu}$, whose conformation is given by a random walk. In this regime, the above form of the free energy (Eq. (2.19)) no longer holds. The correct form of the free energy [57] for $d = 3$ leads to a new force-extension relation [59]:

$$f \sim -\frac{k_B T}{D} \left(\frac{L}{L_0} \right)^{-9/4}. \tag{2.21}$$

2.2.3 Expansion dynamics

The question of how the chain expansion dynamics is affected by confinement has been addressed in several studies [61, 67, 68], since it allows one to measure relaxation times and effective chain spring constants.

Let us then consider the situation where a polymer is much compressed in a cylindrical pore and then allowed to freely expand (see Fig. 2.2(c)). While expanding, the polymer goes through distinct regimes [61]. After strong compression ($L/L_0 \ll 1$), the first expansion will be dominated by excluded volume effects and not so much by the chain properties. In this regime, the chain breaks up into small blobs so that the system essentially behaves like a high-concentration solution of hard-sphere-like monomers. As the chain relaxes further, the blobs grow in size and the nature of the expansion force changes, altering the dynamics from ballistic to subdiffusive [61]. At this point, the blob picture of a linear string of blobs is recovered, and the dynamics is dictated by the interactions between neighbouring blobs. The system reaches an effective adiabatic limit where the polymer goes through quasi-equilibrium states. The

expansion force can then be approximated to an equilibrium force and the force-extension relation takes the form of Eq. (2.20). For larger times, the polymer enters an exponential regime reaching eventually its equilibrium extension L_0.

For the intermediate regime, the time-dependence of the chain extension L can be determined from the equation of motion,

$$f + \frac{1}{2}\,\gamma\,\frac{dL}{dt} = 0, \tag{2.22}$$

where γ is the polymer friction coefficient. Following the approach in [61], this equation can be solved subjected to the initial condition $L(t=0) \equiv L_i$, the extension of the compressed polymer just before it starts expanding. Substituting f by Eq. (2.20), one gets

$$\frac{dL}{dt} = -\frac{2}{\gamma}\,k_{\text{chain}}\,\frac{k_B T}{D}\left[\left(\frac{L}{L_0}\right) - \left(\frac{L}{L_0}\right)^{-2}\right]. \tag{2.23}$$

By defining $\tilde{L} = L/L_0$ and $A = (2k_{chain}k_B T)\,/\,(L_0\gamma D)$, this becomes

$$\frac{d\tilde{L}}{dt} = -A\left(\tilde{L} - \tilde{L}^{-2}\right) \quad \Leftrightarrow \quad \frac{\tilde{L}^2}{\tilde{L}^3 - 1}\frac{d\tilde{L}}{dt} = -A. \tag{2.24}$$

Taking into account that during the polymer expansion $\tilde{L} < 1$, the integral of this equation leads to the solution

$$\ln(1 - \tilde{L}^3) = -3At + \text{const.}$$

$$\Leftrightarrow \;\; \tilde{L}(t) = \frac{L(t)}{L_0} = \left(1 - c\;e^{-3At}\right)^{1/3}, \tag{2.25}$$

where c is a constant defined by the initial condition: $c = 1 - \tilde{L}^3(t=0)$. Therefore,

$$L(t) = L_0\left\{1 - \left[1 - \left(\frac{L_i}{L_0}\right)^3\right]e^{-3At}\right\}^{1/3}. \tag{2.26}$$

According to this solution, there is an initial offset L_i corresponding to the extension of the polymer after being compressed by the piston. For intermediate times, one can expand the solution about zero to get

$$L(t) \approx \left\{\; L_i^3 + 3A\left[L_0^3 - L_i^3\right]t\;\right\}^{1/3} \quad \Rightarrow \quad L(t) \sim t^{\alpha},\; \alpha = 1/3. \tag{2.27}$$

For late times the polymer relaxes exponentially reaching a saturation regime where the extension corresponds to the value at equilibrium in the absence of a compression force: $L(t \to \infty) = L_0$.

In the strong compression regime, where the force-extension relation for the chain is best described by Eq. (2.21), the extension-time scaling relation for intermediate times then becomes

$$L(t) \sim t^{\alpha}, \ \alpha = 4/13. \tag{2.28}$$

2.3 Interaction with proteins

2.3.1 Molecular crowding

In a crowded environment, entropic forces can paradoxically drive the compaction of polymers. This entropic compaction mechanism, known as the depletion attraction [82], is illustrated in Figure 2.3(a). In a solution containing many small particles/crowders[1], compacting the polymer effectively lowers the free energy of the system through an increase of the translational entropy of the crowders [5]. Due to excluded volume effects, the polymer monomers are surrounded by a depletion layer (grey areas in Fig. 2.3(a)) that the centres of mass of the crowders cannot access. When monomers come together, there is partial overlapping of their depletion layers, increasing the volume available to the crowders. In a solution with a lot more small particles than monomers, the polymer chain can therefore be compacted since its monomers will experience depletion forces that compete with their excluded volume interaction. The free energy decrease due to the increased translational entropy is maximal when monomers are in contact, taking the value [5]

$$\Delta\mathcal{F} = -\frac{3}{2}\frac{\sigma}{\sigma_c}\phi_c \ k_B T, \tag{2.29}$$

where σ and σ_c are the sizes of the monomers and crowders, and ϕ_c is the volume fraction of crowders. The decrease in the free energy then grows linearly with the number of crowders. Therefore, for a sufficiently large volume fraction

[1] In the case of the cell nucleus, these represent proteins or nuclear complexes that interact sterically with the DNA.

ϕ_c, the depletion effects become dominant and the polymer chain collapses (see Fig. 2.3(b)).

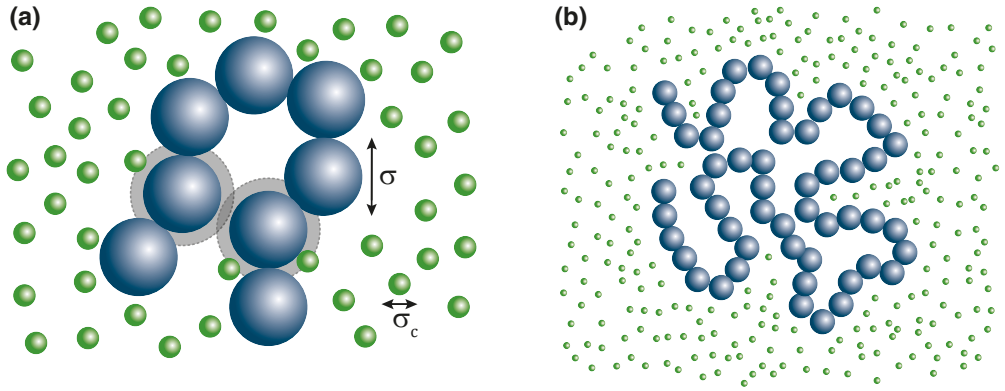

Figure 2.3 *(a) Pictorial representation of the depletion attraction between polymer monomers induced by crowding proteins. Due to excluded-volume interactions, crowders are depleted from the volume surrounding the monomers (grey areas). For a high number of crowders, decreasing the inaccessible volume is entropically favourable, and the monomers are attracted to each other. (b) For a high concentration of proteins, crowding effects can lead to chain collapse.*

Depletion forces play a major role in the compaction of DNA, specially in bacterial cells, where macromolecules are estimated to occupy around $20\% - 30\%$ of the cellular volume [106]. Previous theoretical studies [101] predicted that depletion forces alone can promote phase separation of DNA and cytoplasmic proteins within an *E. coli* cell, leading to the formation of the nucleiod. Indeed, in recent experiments [104], abrupt compaction of bacterial DNA trapped in a nanopore (3-fold compaction) was observed above a volume fraction $\sim 12\% - 17\%$ of molecular depletants. These results were further corroborated by recent simulation and theoretical studies [125] of coarse-grained bacterial chromosomes, which predicted that molecular crowding can induce a continuous phase transition from a swollen to a collapsed globular polymer state.

Molecular crowding also effectively increases the viscosity of the medium [135], hindering the polymer dynamics. For a polymer trapped in a nanopore, one therefore might expect the exponent α in Eq. (2.27) to decrease with the number of crowders.

2.3.2 DNA-binding proteins

In vivo, DNA is not only surrounded by crowding proteins or other organelles, but also by DNA-binding proteins. If the former contribute to a global genome compaction, the latter contribute to a structured folding of DNA. In the polymer picture of DNA, these proteins can be thought of, to a first approximation, as multivalent hard-core spheres that interact attractive and non-specifically with the polymer chain (see Fig. 2.4(a)). This can be the case of transcription factors or other complexes that can bind to multiple sites on the genome simultaneously, forming "molecular bridges" that stabilise DNA loops. When the interaction is strong enough, the bound proteins naturally cluster, through the so-called "bridging-induced attraction" first described in [17], even if no direct protein-protein attractive interaction is considered.

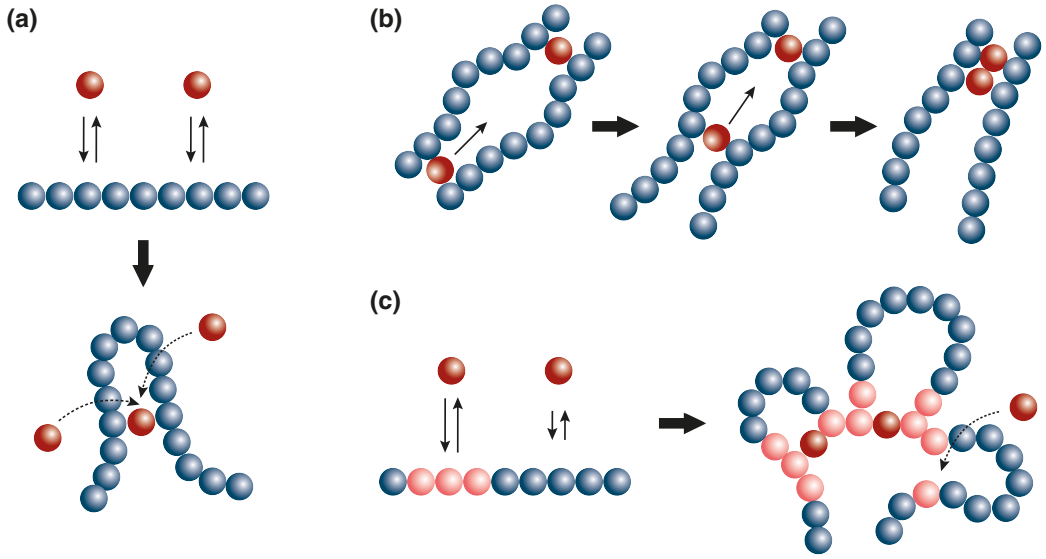

Figure 2.4 *Pictorial representation of the bridging-induced attraction between DNA-binding proteins. (a) Model where multivalent proteins bind non-specifically with the DNA polymer, leading to the formation of molecular bridges between monomers. (b) Multivalent proteins can diffuse along the polymer, and eventually cluster. (c) Model where multivalent proteins bind with stronger affinity with specific sites along the polymer, leading to the formation of polymer loops.*

This clustering is associated with a simple positive feedback loop: multivalent proteins bind to the DNA in multiple places forming molecular bridges; this increases the local concentration of DNA, which in turn facilitates further binding of proteins; the process repeats, and a protein cluster eventually forms. To a first

approximation this model is able to explain the formation of protein clusters in bacteria, like H-NS and DPS clusters [17, 32], or nuclear bodies in eukaryotic nuclei, which are essentially clusters of specialised proteins [16, 83].

Due to the non-specific nature of the polymer-protein interaction, proteins can bind anywhere along the chain and effectively diffuse while binding (see Fig. 2.4(b)). The bridging-induced attraction then leads to protein clusters that grow in size and that end up merging, ultimately yielding one single cluster in steady state [54]. However, this is not a realistic description of protein cluster growth in live cells, where clusters only grow up to a finite size. The model can be easily adapted to correctly account for this arrested growth.

While most DNA-binding proteins in bacteria (e.g. H-NS, HU) and in eukaryotes (e.g. transcription factors) interact non-specifically with DNA and chromatin, i.e. via electrostatic interactions, they also interact specifically with the genome, binding with different affinities to different genomic sequences. This feature can be included in the model by considering proteins that bind with stronger affinity to specific polymer sites (see Fig. 2.4(c)). In this picture, protein clusters still form via the bridging-induced attraction, but do not grow indefinitely, reaching instead a self-limiting size. Since proteins now bridge high affinity sites that are joined by DNA loops, cluster growth involves the creation of more and more loops. While the number of DNA-protein interactions increases linearly with the cluster volume, the entropic cost of creating more DNA loops grows superlinearly with the number of loops. Therefore, cluster growth is arrested when the entropic cost overcomes the enthalpic gain from creating more DNA-protein interactions.

The DNA bridging effect driven by protein binding not only contributes to the genome compaction but also increases the local polymer stiffness, effectively reducing the polymer elasticity. Therefore, for a polymer confined in a nanopore in the presence of such proteins, one expects a slow-down effect of the polymer expansion dynamics, possibly reflected by a decrease of the exponent α in Eq. (2.27).

Molecular Dynamics Simulations

3

In the present Thesis, the elastic, dynamic and conformational properties of both bacterial and eukaryotic chromosomes are studied by means of computational simulations. In particular, simulations are designed using the Molecular Dynamics (MD) method.

First developed in the 50s, MD simulations provide a mechanistic approach to study the physical properties of a system. The system is treated as a collection of interacting particles, whose trajectories (position and velocity) are determined as time evolves. As a result, it is possible to probe time-dependent responses and dynamical properties. Moreover, if the system obeys the ergodic hypothesis (i.e. if every microstate of the system can in principle be reached by the dynamics), then these simulations allow for the calculation of macroscopic thermodynamic properties by simply computing time averages, which then correspond to averages over particle ensembles.

The key aspect of MD simulations is the choice of the right level of coarse-graining of the model. Due to their flexibility, MD methods are easily used to simulate systems based on very high resolution (e.g. atomistic), as well as low resolution models. Since high-resolution simulations incur a high computational cost, it is usual to choose the level of coarse-graining that still retains the key physical elements of the system, but that allows one to reach the longest simulation time scales.

In this chapter I will present the general details of the coarse-grained polymer model and MD simulations that are common to the different systems that will be the subject of study of this Thesis. Later, in the respective chapters, I will give more details specific of each studied case.

3.1 Simulation scheme

The most common Molecular Dynamics simulations consist of determining the trajectories of the system's particles by simply solving the classical equations of motion:

$$m_i \frac{d^2 \boldsymbol{r}_i}{dt^2} = \boldsymbol{F}_i = -\boldsymbol{\nabla}_i V \ , \tag{3.1}$$

where m_i and $\boldsymbol{r}_i$ are the mass and position of particle i, respectively. $\boldsymbol{F}_i$ is the resultant force exerted on particle i, which can be determined given the potential energy V. The latter results from the interactions with all the other particles, with the system's boundaries (e.g. reflective walls), and/or with external fields. Therefore, in order to predict the behaviour of the system, one needs to define the potentials describing the relevant interactions in a specific case. These are of two types: bonded and non-bonded potentials. Bonded potentials describe interactions between particles connected at a molecular level, like successive monomers in a polymer, whereas non-bonded potentials describe attractive and repulsive interactions between non-connected particles, such as excluded-volume or DNA-protein interactions.

3.1.1 Bonded potentials

The connectivity between neighbouring polymer monomers, of diameter σ, is well described by the Finitely Extensible Non-linear Elastic (FENE) potential:

$$U_{\mathrm{FENE}}(r_{i,i+1}) = -\frac{K_{\mathrm{FENE}}}{2} r_0^2 \ln\left[1 - \left(\frac{r_{i,i+1}}{r_0}\right)^2\right] \tag{3.2}$$

for $r_{i,i+1} < r_0$, and $U_{\mathrm{FENE}}(r_{i,i+1}) = \infty$ otherwise. Here, $r_0 = 1.6\sigma$ is basically the maximal extent of the bond, $r_{i,i+1}$ the distance between the ith monomer and its nearest neighbour (the $(i+1)$th) along the chain, and $K_{\mathrm{FENE}} = 30 k_B T/\sigma^2$ the bond energy. This choice of parameters ensures that polymer segments cannot cross each other, correctly describing inter-segment interactions.

Another useful way to describe bonds between monomers is through the harmonic potential

$$U_{\mathrm{harm}}(r_{i,i+1}) = \frac{K_{\mathrm{harm}}}{2} \left[r_{i,i+1} - r_0\right]^2 , \tag{3.3}$$

i.e., bonds are modelled as harmonic springs of stiffness K_{harm} and equilibrium length r_0. Although this potential does not prevent adjacent monomers from overlapping, it is convenient, in a pre-equilibration step, when relaxing a polymer from an initial conformation where the chain bonds can be too extended. Since this potential continually increases with increasing $r_{i,i+1}$, it will yield a soft attraction between connected monomers that are far away without generating numerical divergences, as in the case of the FENE potential.

The harmonic potential is also used to generally model bonds between particles other than consecutive monomers (chain connectivity), such as bonds between non-neighbouring monomers, or between monomers and other particles.

The bending rigidity of the polymer is set through the Kratky-Porod potential for every three adjacent monomers

$$U_{\text{bend}}(r_{i,i+1}, r_{i+1,i+2}) = K_{\text{bend}} \left(1 + \frac{\boldsymbol{r}_{i,i+1} \cdot \boldsymbol{r}_{i+1,i+2}}{r_{i,i+1} r_{i+1,i+2}}\right)$$
$$= K_{\text{bend}}(1 + \cos(\theta)), \tag{3.4}$$

where θ is the angle between three consecutive monomers, and K_{bend} the bending energy. K_{bend} sets the flexibility of the polymer since this determines the persistence length: $l_p = (K_{\text{bend}}/k_B T)\ \sigma$. Two values to keep in mind are the persistence length of hydrated B-DNA, $l_p \approx 50$ nm$= 20\sigma_{\text{DNA}}$ [47, 106], and of the 30-nm chromatin fibre, $l_p \approx 90$ nm$= 3\sigma_{\text{chrom}}$ [6, 73], where $\sigma_{\text{DNA}} = 2.5$ nm and $\sigma_{\text{chrom}} = 30$ nm are the monomer size or, in other words, the polymer thickness of DNA and chromatin.

3.1.2 Non-bonded potentials

Every particle interacts via a non-bonded potential with all the other particles and with the system's boundary wall (if existent). It is common to simply consider 1-body and 2-body interaction potentials, as long as these reproduce the essential physics of the system.

Particle-particle excluded volume interactions are well described by the Weeks-Chandler-Andersen potential (WCA) [139], which is essentially a shifted truncated form of the usual Lennard-Jones (LJ) potential that yields a purely

repulsive interaction:

$$U_{\mathrm{WCA}}(r_{ij}) = 4k_BT\left[\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - \left(\frac{d_{ij}}{r_{ij}}\right)^{6} + \frac{1}{4}\right]\Theta(2^{1/6}d_{ij} - r_{ij}). \tag{3.5}$$

Here, r_{ij} is the distance between the ith and jth particles, d_{ij} the minimum distance between particles (for two particles of size σ: $d_{ij} = \sigma$), and Θ the usual Heaviside function (1 for $r_{ij} < 2^{1/6}d_{ij}$, and 0 otherwise). According to this shifted form of the potential, $U_{\mathrm{WCA}} = 0$ at the cut-off distance $2^{1/6}d_{ij}$, meaning that long-range interactions can be discarded without introducing discontinuities.

Another way of describing excluded volume interactions is through a "soft" potential given by

$$U_{\mathrm{soft}}(r_{ij}) = K_{\mathrm{soft}}\left[1 + \cos\left(\frac{\pi r_{ij}}{r_c}\right)\right]\Theta(r_c - r_{ij}), \tag{3.6}$$

where K_{soft} is the height of the potential at $r_{ij} = 0$, and r_c the cut-off distance. This potential is useful in a pre-equilibration step, when relaxing the system from an unrealistic initial conformation with partially overlapping particles. Since it does not diverge when $r_{ij} \to 0$, it softly pushes particles apart.

Proteins, modelled as spherical beads, can interact sterically (crowders) or attractively (multivalent DNA-binding proteins) with the polymer. The steric interaction is given by the WCA potential in Eq. (3.5), whereas the attractive protein-monomer interaction is given by the LJ potential

$$U_{\mathrm{LJ\ shift}}(r_{ij}) = \left[U_{\mathrm{LJ}}(r_{ij}) - U_{\mathrm{LJ}}(r_c)\right]\Theta(r_c - r_{ij}), \tag{3.7}$$

$$\text{where} \qquad U_{\mathrm{LJ}}(r) = 4\epsilon\left[\left(\frac{d_{ij}}{r}\right)^{12} - \left(\frac{d_{ij}}{r}\right)^{6}\right]. \tag{3.8}$$

The potential is shifted so that it goes to zero at the cut-off distance to avoid discontinuities. r_c and ϵ vary depending on the protein-monomer interaction affinity (see Chapters 4, 5, and 6 for details specific to the different systems).

Depending on the problem considered, both proteins and the polymer can interact either sterically or attractively with the system's confining walls. In the case of a hard wall, the steric and attractive particle-wall interactions are defined by the 1-body form of the WCA potential in Eq. (3.5) and of the LJ potential in Eq. (3.7), respectively, with $d_{ij} \to \sigma$ and $r_{ij} \to r_w$ (the distance from the particle

to the wall). For particles that move beyond the wall, $U_{\text{wall}} = \infty$.

It can be also useful to model soft walls, where particles are allowed to move through the wall for a short time before they are pulled back again. The particle-wall interaction is then defined by considering that, when particles go beyond the wall to the "forbidden" region, they will be subjected to a force that pulls them back to the wall. This can be modelled by the indentation potential, which yields a stronger particle-wall interaction than the usual harmonic potential:

$$U_{\text{indent}}(r_w) = \frac{K_{\text{indent}}}{3} r_w^3 \tag{3.9}$$

for particles beyond the wall, and $U_{\text{indent}} = 0$ otherwise. Here r_w is the distance from the particle to the wall on the "forbidden" region and K_{indent} sets the wall strength. Basically, according to this model, particles only "feel" the wall when they go past it. This potential is particularly useful when modelling a moving wall, since this can go through the particles without generating divergencies.

3.2 Brownian Dynamics

In the cell, DNA and proteins are effectively dispersed in a viscous solvent (water plus many small salt molecules), and therefore move at low Reynolds numbers. In this regime, where the hydrodynamic flow of the solvent caused by the motion of the constituents can be disregarded, it is usual to model the solvent implicitly. Instead of integrating the deterministic equations of motion for all particles (including solvent molecules), the polymer monomers and proteins can simply be coupled with a bath at fixed temperature. This approach highly reduces the computational cost of the simulations allowing to evolve the system for long time scales. The interaction with the solvent is now modelled by considering that each particle is subjected to a force representing the many collisions with the small solvent molecules, which cause the particle to undergo Brownian (random) motion. Now the motion of monomers and proteins is no longer deterministic, but has a stochastic nature. All particles will also experience a drag force due to the viscosity of the medium, which at low Reynolds numbers is proportional to their velocity. The motion of each particle is then described by the Langevin equation

$$m\frac{d^2\boldsymbol{r}_i}{dt^2} = -\boldsymbol{\nabla}_i V - \gamma_i \frac{d\boldsymbol{r}_i}{dt} + \sqrt{2k_B T \gamma_i}\boldsymbol{\eta}_i(t), \tag{3.10}$$

where $\gamma = m/\tau_{\text{in}}$ is the friction due to the solvent, with τ_{in} being the inertial time (i.e., the velocity decorrelation time), and $\boldsymbol{\eta}_i(t)$ is a vector representing random uncorrelated noise, such that

$$\begin{aligned} \langle \eta_{i\alpha}(t) \rangle &= 0, \\ \langle \eta_{i\alpha}(t)\eta_{j\beta}(t') \rangle &= \delta_{\alpha\beta}\delta(t-t')\delta_{ij}, \end{aligned} \tag{3.11}$$

where α and β indicate Cartesian components, δ_{ij} is Kronecker's delta, and $\delta(t-t')$ denotes Dirac's delta function. For observation times much larger than τ_{in}, inertial forces become negligible and the motion of particles is effectively described by the "over-damped" form of Eq. (3.10):

$$\frac{d\boldsymbol{r}_i}{dt} = -\frac{1}{\gamma_i}\boldsymbol{\nabla}_i V + \sqrt{2D_{\text{diff},i}}\boldsymbol{\eta}_i(t), \tag{3.12}$$

where $D_{\text{diff},i}$ is the diffusion coefficient of particle i given by the Einstein relation $D_{\text{diff},i} = k_B T/\gamma_i$.

3.2.1 Time integration of the equations of motion

Eq. (3.12) is evolved using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) (http://lammps.sandia.gov). Integration is performed in the canonical ensemble, where the number of particles N, volume V and temperature T of the system are preserved throughout the simulation. The temperature is kept constant by a Langevin thermostat, which ensures thermal equilibrium between the system of particles and a heat bath by regulating the magnitude of the thermal noise via the fluctuation-dissipation relation $D_{\text{diff}} = k_B T/\gamma$.

The numerical integration of the equations of motion is performed using the velocity Verlet algorithm [132] for a constant time step $\Delta t = 0.01\tau$, where τ is the simulation time unit. Verlet-type algorithms are very useful since they provide good numerical stability and ensure energy conservation. Even though the energy of the system is not conserved during the simulation due to the coupling of the particles to the thermal bath, this algorithm ensures that once the system reaches equilibrium the energy is (on average) conserved. The basic Verlet algorithm can be derived by simply considering the Taylor expansions of the position vector $\boldsymbol{r}$

around time t:

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{r}'(t)\Delta t + \frac{1}{2}\boldsymbol{r}''(t)\Delta t^2 + \frac{1}{6}\boldsymbol{r}'''(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \tag{3.13}$$

$$\boldsymbol{r}(t-\Delta t) = \boldsymbol{r}(t) - \boldsymbol{r}'(t)\Delta t + \frac{1}{2}\boldsymbol{r}''(t)\Delta t^2 - \frac{1}{6}\boldsymbol{r}'''(t)\Delta t^3 + \mathcal{O}(\Delta t^4). \tag{3.14}$$

Here Lagrange's notation for differentiation is used (i.e., $\boldsymbol{r}' \equiv d\boldsymbol{r}/dt$, etc.) and thus $\boldsymbol{r}'(t) \equiv \boldsymbol{v}(t)$, the velocity, and $\boldsymbol{r}''(t) \equiv \boldsymbol{a}(t) = \boldsymbol{F}(t)/m$, the acceleration of the particle. Adding the two equalities together, one gets the standard Verlet equation to get the position vector for the next integration time step:

$$\boldsymbol{r}(t+\Delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\Delta t) + \frac{\boldsymbol{F}(t)}{m}\Delta t^2 + \mathcal{O}(\Delta t^4), \tag{3.15}$$

where $\boldsymbol{F}(t) = -dU(t)/d\boldsymbol{r}$. The equation for the velocity can be obtained using the same approach, but now by subtracting the two Taylor expansions:

$$\boldsymbol{v}(t) = \frac{1}{2\Delta t}\left[\boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t-\Delta t)\right] + \mathcal{O}(\Delta t^2). \tag{3.16}$$

The new position is accurate up to fourth order in Δt, whereas the new velocity is only accurate up to second order. A disadvantage of this method is that the calculation of the next position vector requires the storing of information for the two previous time steps.

The velocity Verlet algorithm circumvents this problem by using the current velocity to calculate the next position. This algorithm is a reformulation of the standard Verlet, where, by manipulating Eqs. (3.13)–(3.15), it yields equations mathematically equivalent to Eq. (3.15) and Eq. (3.16), but numerically more convenient. To derive the velocity Verlet relations it is useful to define the variable

$$\boldsymbol{z}(t) = \frac{1}{\Delta t}\left[\boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t)\right]. \tag{3.17}$$

Eq. (3.15) and Eq. (3.16) can be, respectively rewritten as:

$$\boldsymbol{z}(t) - \boldsymbol{z}(t-\Delta t) = \frac{\Delta t}{m}\boldsymbol{F}(t) \tag{3.18}$$

$$\boldsymbol{v}(t) = \frac{1}{2}\left[\boldsymbol{z}(t) + \boldsymbol{z}(t-\Delta t)\right]. \tag{3.19}$$

By manipulating Eq. (3.19) and using the equality in Eq. (3.18), one obtains the velocity Verlet equations to get the position and velocity for the next time step:

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\Delta t + \frac{1}{2m}\boldsymbol{F}(t)\Delta t^2 \tag{3.20}$$

$$\boldsymbol{v}(t+\Delta t) = \boldsymbol{v}(t) + \frac{\Delta t}{2m}\left[\boldsymbol{F}(t+\Delta t) + \boldsymbol{F}(t)\right]. \tag{3.21}$$

These now only depend on the current position and velocity, while keeping the numerical precision of the standard Verlet equations.

3.2.2 Mapping from system to physical units

When performing MD simulations, it is convenient to consider a set of fundamental units given by the systems' fundamental quantities m and σ (mass and size of a polymer monomer), and $\epsilon = k_B T$. All other quantities are simply given as multiples of these. Therefore, in this Thesis, the properties of the system will be reported in terms of these fundamental units. Specifically, masses will be given in units of m, distances in units of σ, energies in units of $k_B T$, forces in units of $k_B T/\sigma$, and time in units of the Lennard-Jones (LJ) time $\tau \equiv \tau_{LJ} = \sigma\sqrt{m/k_B T}$. The LJ time scale gives the time taken for a particle of mass m to be repelled a distance σ by a force $k_B T/\sigma$. In other words, it gives the time scale for collisions between particles interacting via a LJ potential.

In the system there are two further time scales with an intuitive physical meaning: the inertial time $\tau_{\text{in}} = m/\gamma$, introduced above, and the Brownian time $\tau_B = \sigma^2/D_{\text{diff}}$. The former is tuned as a simulation parameter, and is usually set to either $\tau_{\text{in}} = \tau/2$ or $\tau_{\text{in}} = \tau$ – this choice ensures that the over-damped limit in Eq. (3.12) becomes a good approximation of Eq. (3.10) after only a few time steps. The latter can be understood as the time taken for a particle to diffuse across its own size, σ. The Brownian time is set by the friction γ through the Einstein's relation $D_{\text{diff}} = k_B T/\gamma$. Under the usual choices for τ_{in}, γ is usually set to either $\gamma = 2m/\tau$ or $\gamma = m/\tau$, which leads, respectively, to a Brownian time scale $\tau_B = 2\tau = 4\tau_{\text{in}}$ or $\tau_B = \tau = \tau_{\text{in}}$.

The system's units can be easily mapped to real units by fixing the fundamental quantities. The temperature is usually kept constant at a value $T = 300$ K, and therefore $\epsilon \sim 0.6$ kcal/mol, however m and σ will depend on the system

to be studied. This Thesis will focus on the properties of DNA at two levels of compaction: dsDNA and chromatin. For dsDNA, σ corresponds to the thickness of the double helix, which for hydrated B-DNA is $\sim$ 2.5 nm (at physiological concentrations of $\sim$ 150 mM). Given that for B-DNA the distance between consecutive base pairs is 0.34 nm, each polymer monomer represents $\sim$ 7.4 bp, and thus m is the mass of a 7.4 bp long DNA segment. For chromatin, σ corresponds to the thickness of the fibre ($\sigma = 30$ nm). The typical level of DNA compaction in a 30-nm chromatin fibre is $\sim$ 100 bp/nm [115], hence each monomer represents $\sim$ 3 kbp, with m being the mass of a chromatin segment equivalent to 3 kbp. Note that these choices lead to a large mass with respect to realistic values. However this is not significant to the studies proposed in this Thesis, since the value of the mass will only affect the dynamics for times smaller than the Brownian time, and we are interested in analysing the behaviour of the system for times much larger than the Brownian time.

By using this mapping for the mass, length and energy, all of the other physical quantities can be easily mapped to real units. For instance, in the case of dsDNA, one simulation unit of force, k_BT/σ, corresponds to 1.64 pN.

The mapping of the simulation time scales to real units needs to be done more carefully. Since we are interested in the long time over-damped behaviour of the polymer, and not in resolving the fine details of the inertial collisions, a common approach consists of mapping the simulation time by considering the Brownian time scale instead of the LJ one. By using the Stokes' formula for the friction of beads of size σ in a solvent of viscosity ν

$$\gamma = 3\pi\nu\sigma, \tag{3.22}$$

and by using the viscosity appropriate for aqueous solvents $\nu = 1$ cP, the Brownian time scale is simply given by

$$\tau_B = \frac{3\pi\nu\sigma^3}{k_BT} = \begin{cases} 36 \text{ ns} & \text{for } \sigma = 2.5 \text{ nm}, \\ 62\ \mu\text{s} & \text{for } \sigma = 30 \text{ nm}. \end{cases} \tag{3.23}$$

The simulation time is finally mapped either through $\tau = \tau_B/2$ or $\tau = \tau_B$, depending on the value chosen for τ_{in} (the exact details are given in the Chapters devoted to the specific systems).

The choice $\nu = 1$ cP for the nucleoplasm is, however, an underestimate, specially for 30 nm beads. Another useful way of mapping the Brownian time to real units, which avoids making assumptions on the solvent viscosity, is by matching the diffusional properties of the polymer and of real DNA/chromatin. The mean square displacement (MSD) averaged over all polymer beads is used to fit the simulation time unit to the experimentally measured MSD of fluorescently labelled chromosomal loci. The MSD describes the average displacement of a particle after a given time lag τ_{lag}:

$$MSD(\tau_{\text{lag}}) = \langle (\boldsymbol{r}(t + \tau_{\text{lag}}) - \boldsymbol{r}(t))^2 \rangle \tag{3.24}$$

where here $\langle \cdots \rangle$ stands for the average over time. For particles undergoing Brownian motion, the MSD depends linearly on τ_{lag} as:

$$MSD(\tau_{\text{lag}}) = 6 D_{\text{diff}} \tau_{\text{lag}}. \tag{3.25}$$

By matching the averaged simulation and experimental values for D_{diff}, τ can be directly mapped to real units. The advantage of this mapping approach is that it depends on the system's particle concentration, and therefore is more reliable when comparing dilute simulations with experiments performed at physiological concentrations. This is the case of the simulations performed in Chapter 5, where chromatin is modelled at low concentrations to avoid non-physical confinement effects (see Chapter 5 for exact details on mapping time units).

Entropic elasticity and dynamics of the bacterial chromosome

4

As reviewed in Chapter 1, the genome of living bacteria is under remarkable confinement [1, 86]. For instance, the chromosome of *E. coli* would be over 1 mm long if stretched out, yet it needs to fit within the bacterial cell which is a $2 \times 1 \times 1$ μm ellipsoid.

There are at least four mechanisms responsible for the compaction of bacterial chromosomes. First, and most obviously, the genome is confined within the cell wall. However it does not occupy the entire cell, but instead it is found to be compacted such that all or most of the genome occupies only a cellular sub-volume called nucleoid. Since the nucleoid is not delimited by a nuclear membrane, the bacterial cell must rely on other mechanisms to further compact the chromosomes. Second, there is a depletion attraction between genome segments induced by the crowding of non-DNA-binding macromolecules [35] (see Sec. 2.3.1). Third, the genome is associated with a number of architectural or nucleoid associated proteins (NAPs) [19, 128] which can bind the DNA at more than one point, creating effective DNA-DNA attractive interactions which help to reduce the space occupied by the chromosome (see Sec. 2.3.2). Fourth, the bacterial chromosome is in fact negatively supercoiled (with a supercoiling degree of $\sim -5\%$) [8]. *In vivo*, bacterial DNA is slightly underwound, so that in a length of the double helix, which would have 20 turns at thermodynamic equilibrium, there are only 19 turns. The torsional stress arising from this twist deficit can be relaxed by negatively writhing the double helix. This then creates a local folding of the DNA, again favouring compaction.

As a first approximation, one can simply consider the most basic compaction

mechanism: view the bacterial chromosome as a biopolymer under tight confinement. The beginning of this century has seen a number of both simulation [58, 59, 61, 62] and experimental [104, 112] studies that probe the effect of confinement on the elastic, dynamic and organisational properties of DNA. According to this tight confinement model [60], there is a large decrease in entropy when the chromosome is confined in the cell, which yields an entropic pressure, or force, on the confining membrane. This naturally explains why the bacterial chromosome tends to expand when the confining cell wall is removed.

The entropic force exerted by a confined chromosome was measured in an interesting experiment by Pelletier *et al.* [104]. In that work, single bacterial cells were trapped in cylindrical nanopores, with diameter just larger than the width of the bacteria, and the height much larger than its length. The cells were then lysed, freeing the enclosed DNA. This expanded along the nanopore, increasing its conformational entropy. The DNA was found to reach a height about ten times larger than that of the bacterium. Using a micro-piston, consisting of a colloidal bead controlled by means of optical tweezers, the bacterial chromosome was subsequently compressed and the entropic force exerted on the piston was measured. The study reported entropic forces much larger than expected on the basis of a simple theory (reviewed in Section 2.1.1) for Gaussian – i.e., infinitely thin – polymers, with the known persistence length of DNA, 50 nm.

Inspired by the work just described [104], in this Chapter I shall present a simulation study of a system closely related to the experiments, with the aim of quantifying the contribution of excluded volume and polymer-protein interactions on the entropic compression elasticity of the bacterial chromosome and its expansion dynamics. More specifically, while previous theoretical works [10, 58, 59, 61, 62, 64, 69, 75, 123, 125, 140] focused on the role of polymer-polymer interactions and confinement on the chromosome properties, here I explicitly quantify the effect of bacterial proteins, whether binding to DNA or not.

I will, therefore, consider a bacterial chromosome, modelled for simplicity as a linear polymer, confined in a thin cylinder and subjected to the action of a piston, i.e. a spherical bead that moves under an applied external force. This set-up allows the measurement of the force exerted on the piston as a function of DNA compression, as well as the polymer expansion dynamics upon removal of the piston.

Although in *E. coli* DNA is actually a circular molecule, the scaling theory and previous work suggests that the main contribution to the polymer response is given by polymer confinement (entropy and free energy loss). Hence one can expect the polymer topology to have only a small effect – the results described in Ref. [105] for simulations with circular DNA further support this view.

The simulation results in Sections 4.2 and 4.3 show that steric effects and the presence of proteins strongly affect both the entropic elasticity of the DNA under compression, and also the dynamics of its extension once the piston is removed. In particular, non-DNA-binding proteins exert an osmotic pressure which can be at least as large as the entropic force exerted by the DNA, and for a sufficiently high number of proteins the DNA entropic response is effectively dwarfed, leaving an undetectable signature in the force-compression curves. The macromolecular crowding proteins introduce is also important in determining the polymer dynamics: DNA expansion is significantly hampered.

In Section 4.3.3, I also present an interesting "popping-off" kinetics displayed by DNA-binding proteins that emerges upon tuning the thermodynamic interaction with the chromosome. Specifically, for a range of DNA-protein binding affinities, there is stable protein binding under confinement, but not in solution.

4.1 Modelling a single chromosome within a nanopore

4.1.1 The system

In light of the experiments performed by Pelletier *et al.* [104], in this study the elastic and dynamic properties of a single bacterial chromosome are probed by considering a DNA segment trapped in a cylindrical pore, that can be compressed by a piston in the absence or presence of proteins. For simplicity, explicit electrostatic interactions between proteins and/or DNA are not considered, even though these are charged in reality. This approximation is taken on the basis that for physiological concentrations of a monovalent salt ($\sim$ 150 mM), the Debye length is $\sim$ 1 nm [3], which is below the particle sizes considered here. Therefore, in practice, electrostatic interactions are heavily screened. The results described in Appendix A, for simulations where the particles' charge is taken into account,

essentially confirm that one can rely on this simplification.

The DNA segment is modelled as a linear self-avoiding polymer composed of spherical monomers with diameter $\sigma \equiv 2.5$ nm, according to the details previously described in Chapter 3. Proteins are modelled, for simplicity, as spherical beads with diameter σ [1]. These interact sterically with each other through the WCA potential in Eq. (3.5). Two types of proteins are considered: non-DNA-binding and DNA-binding proteins. The former represent molecular crowders, i.e., proteins that do not bind to DNA, and thus are modelled as beads that interact sterically with the DNA. The latter represent bacterial histone-like proteins (H-NS) [32], or DNA-binding proteins that are activated following bacterial starvation (DPS) [141], which, to a first approximation, bind DNA nonspecifically, and thus are modelled as multivalent beads that interact attractively with any DNA monomer through the LJ potential in Eq. (3.7). For the simulations where both types of proteins are present, the parameters for the attractive LJ potential are $r_c = 3\sigma$ and $\epsilon = 2.5k_BT$, yielding a moderate attraction. For the simulations where only the DNA-binding proteins are present, $r_c = 1.5\sigma$ and ϵ can take values in the range $[2, 4]$ to explore the effect of the interaction strength.

The piston is modelled as a rigid sphere whose diameter is slightly larger than the confining cylindrical pore's diameter, so as to avoid particles from escaping around the sides. The interaction between the piston and the confined particles is purely repulsive and described by the WCA potential in Eq. (3.5). The particle-cylinder wall interactions are described by the potential in Eq. (3.9).

The number of particles in the system is chosen so that the ratio of the number of proteins to the number of DNA beads is that found *in vivo*. In *E. coli* cells, the chromosome consists of $\sim 4.6 \times 10^6$ bp, which in our model corresponds to a polymer with $\sim 5 \times 10^5$ beads. The total number of proteins is estimated to be $M \sim 10^6$, from which around 3% are DNA-binding. This corresponds to a ratio of ~ 2 proteins per DNA bead. Since the computational cost associated with a 5×10^5 bead polymer is extremely high, we instead consider a smaller system with $N = 1000$ beads. We then explore the influence of proteins in the system, by varying the protein number up to $M = 2000$, which represents the true protein-DNA ratio. For simulations where both crowders and DNA-binding

[1] In Appendix B, I present selected results where proteins of diameter 2σ are considered, which matches more closely the size of a typical bacterial protein [106].

proteins are present, the number of DNA-binding proteins is $0.03M$.

4.1.2 Simulation details

The DNA polymer is initialised as a chain of monomer locations that follow an ideal random walk trajectory, confined in a cylindrical pore (see Fig. 4.1(a)). It is then allowed to relax into a semi-flexible self-avoiding polymer by performing an equilibration run, for $5 \times 10^4 \tau$, where consecutive polymer beads are connected via a harmonic potential and the excluded volume interactions are set by a soft potential:

$$U_{\text{eq}} = \sum_{i=1}^{N-1} U_{\text{harm}}(r_{i,i+1}) + \sum_{i=1}^{N-2} U_{\text{bend}}(r_{i,i+1}, r_{i+1,i+2}) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} U_{\text{soft}}(r_{ij}) + \sum_{i=1}^{N} U_{\text{indent}}(r_{w,i}). \tag{4.1}$$

The persistence length is set from the beginning to the realistic value $l_p = 20\sigma$, and the harmonic and soft potential parameters are chosen so as to avoid chain crossings: $K_{\text{harm}} = K_{\text{soft}} = 100 k_B T$.

The system is considered to be equilibrated when the polymer extension L, measured as the largest distance between two DNA beads along the confining cylinder's axis, no longer increases on average (see Fig. 4.1(b)).

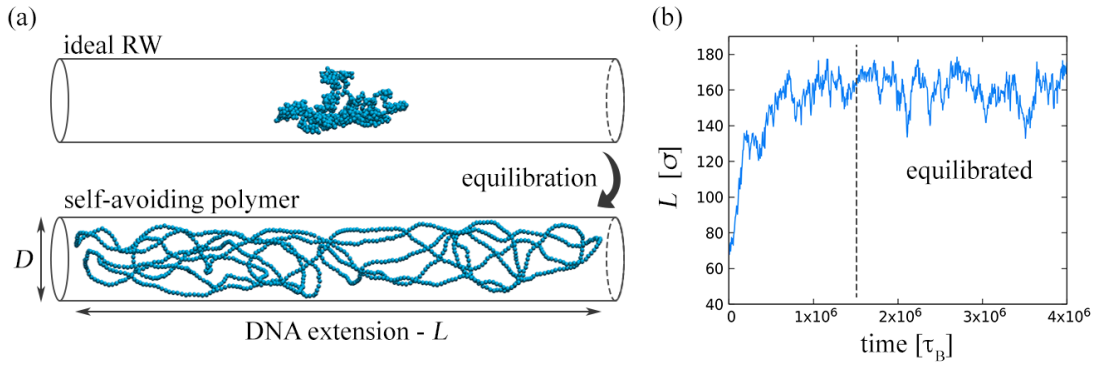

Figure 4.1 *(a) Diagram of the model for a DNA polymer, confined in a cylindrical channel, equilibrating from the initial configuration. The polymer is initialised as an ideal random walk and then let to relax into a semi-flexible self-avoiding polymer. (b) Time evolution of the polymer extension L during equilibration. The system is considered to reach equilibrium when L plateaus (see dashed line).*

In this study we considered cylindrical pores of two different diameters: $D = 20\sigma$ and $D = 40\sigma$ (the former case is shown in Fig. 4.1 and the latter in Fig. 4.2). The results for the two diameters are very similar, hence in this Thesis I will only present and discuss the results for the case where $D = 40\sigma$.

In order to verify if the chosen cylinder dimensions render realistic bacterial conditions, we compared the radius of gyration, R_g, of the unconfined DNA (i.e., in the absence of the cylinder) and the relaxed extension of the confined DNA in the absence of the piston L_0. The radius of gyration was estimated to be

$$R_g^2 = \frac{1}{2N^2} \sum_{i,j} [\boldsymbol{r}_i - \boldsymbol{r}_j]^2 \Rightarrow R_g \sim 62\sigma, \tag{4.2}$$

while its relaxed extension L_0 was measured as $L_0 \sim 152\sigma$. Therefore, the system is in the semi-dilute regime which is realistic for the bacterial chromosome.

For the DNA-protein simulations, we considered two situations. In the first (Fig. 4.2(a)), we only included crowding proteins (green spheres), which are the majority *in vivo* (in *E. coli* there are about 10^6 proteins, of which only 3% are estimated to be DNA-binding). In the second (Fig. 4.2(b)), we included a fraction of DNA-binding proteins (magenta spheres). In both situations, the system is initialised with the proteins randomly distributed inside the confining cylinder. An equilibration run is performed where excluded volume interactions are, again, set by a soft potential, and attractive DNA-protein interactions are set by the LJ potential in Eq. (3.7). The snapshots in Figure 4.2 correspond to equilibrated conformations.

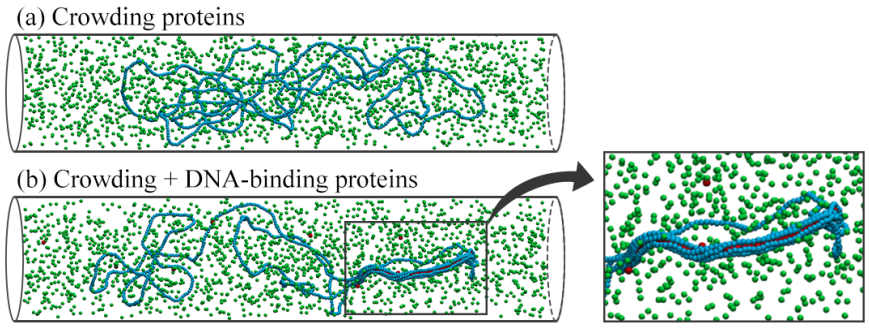

Figure 4.2 *Diagram of the model for a relaxed DNA polymer in the presence of (a) crowding proteins (green spheres) and (b) DNA-binding proteins (magenta spheres). DNA-binding proteins naturally cluster, inducing a local tubular folding of DNA (inset).*

While we include no direct protein-protein attractive interaction, DNA-binding

proteins naturally cluster through the so-called "bridging-induced attraction", as described in Chapter 2 (see inset in Fig. 4.2(b)). For the protein size considered here, this attraction leads to the formation of elongated protein clusters associated with a local tubular folding of DNA where the protein-associated segments are parallel to each other. Ref. [17] shows that larger proteins lead to DNA wrapping around them. However this is not realistic for H-NS which is thought to form linear clusters [17, 32], as in Figure 4.2.

In summary, in this study we performed simulations as follows. First, we considered compression simulations with modified protein density. Second, we simulated the dynamics of entropic expansion, by first compressing the DNA with a large force, allowing the system to reach equilibrium, and subsequently removing the piston. Third, we have varied the strength of the interaction between DNA-binding proteins and DNA.

Brownian Dynamics simulations were performed as described in Chapter 3. The friction due to the solvent was set to $\gamma = 2m/\tau$, which leads to a Brownian timescale $\tau_B = 2\tau = 4\tau_{\text{in}}$, where $\tau_B = 36$ ns (see Sec. 3.2.2 for more details). Simulations were run for a total of 3×10^4 simulation time units (τ) or more.

Table 4.1 summarises the simulation parameters.

$\sigma_{\text{DNA}} = \sigma_{\text{prot}} = 2.5$ nm	Proteins
$l_p = 20\sigma = 50$ nm	DNA : crowders interaction – WCA
$N_{\text{DNA}} = 1000$ beads	DNA : DB-prot interaction – LJ
$M_{\text{crowders}} = [0, 2000]$ proteins	with crowders – $\epsilon = 2.5k_BT$
$M_{\text{DB-prot}} = 0.03M_{\text{crowders}}$	$r_c = 3\sigma$
$D_{\text{cylinder}} = 40\sigma$	without crowders – $\epsilon = [2, 4]k_BT$
$\gamma_{\text{solvent}} = 2m/\tau$	$r_c = 3\sigma$

Table 4.1 *Simulation parameters for the bacterial chromosome model.*

4.2 Chromosome compression beyond the entropic spring regime

The DNA compression experiments performed in Ref. [104] were simulated here by applying a uniform force to the piston bead, which in turn compresses

the DNA polymer against the cylinder walls (Fig. 4.3(a)). For each value of the applied force, the DNA extension was recorded after equilibration. The system is considered to reach equilibrium once the normalised DNA extension L/L_0 plateaus (Fig. 4.3(b)).

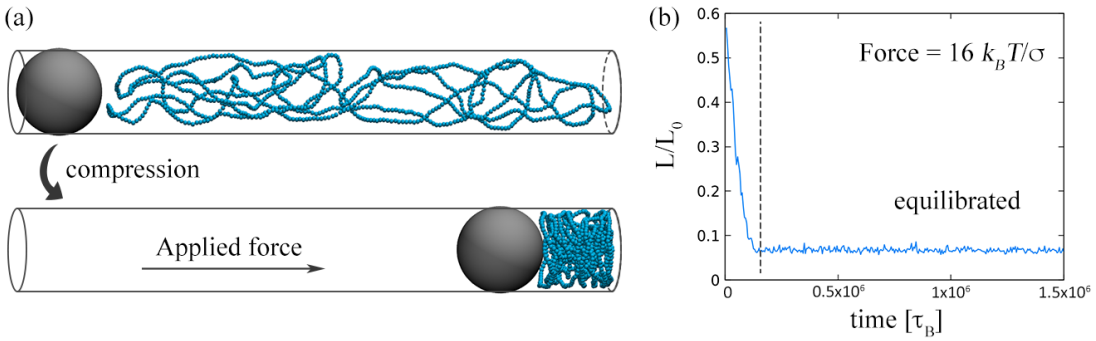

Figure 4.3 *(a) Diagram of the model for a DNA polymer being compressed by a piston subjected to an external force of* $16k_BT/\sigma$*. (b) Time evolution of the normalised DNA extension* L/L_0 *during compression. For each value of the force, the DNA extension is recorded after equilibration (see dashed line), by averaging over several time units.*

Figure 4.4(a) summarises the resulting force-extension curves from simulations (blue circles) and the entropic spring theory (full line). The latter is obtained by fitting Eq. (2.20) to the numerical results, through the spring constant k_{chain}, in the weak compression regime. There are two different regimes. First, for $L/L_0 > 0.1$ the entropic spring theory reviewed in Section 2.2.2 (which is also the theory used in Ref. [104]) works well, and there is a very sensitive dependence of the DNA extension on the compression force: the extension reduces to 10% of its relaxed value, L_0, for a force of just ~ 3 simulation units (corresponding to about 5 pN – see mapping in Sec. 3.2.2). Second, for $L/L_0 < 0.1$, there is a sharp deviation from the theory: the numerical estimate of the entropic force is more than an order of magnitude larger than predicted. This deviation originates from the assumption made in the scaling theory that a polymer confined in a cylindrical pore effectively behaves as a linear chain of polymer blobs of size D. This assumption does not hold in the tightly confined regime, where DNA segments are forced into close contact and the polymer "breaks" into smaller blobs, so that steric interactions dominate over the entropic response. Figure 4.4(b) further suggests that close packing and many body interactions are the main contributions to this strong steric repulsion in the tightly confined regime, as it can be seen from the fact that the volume fraction reaches a large value (~ 0.3) for the largest compression

force.

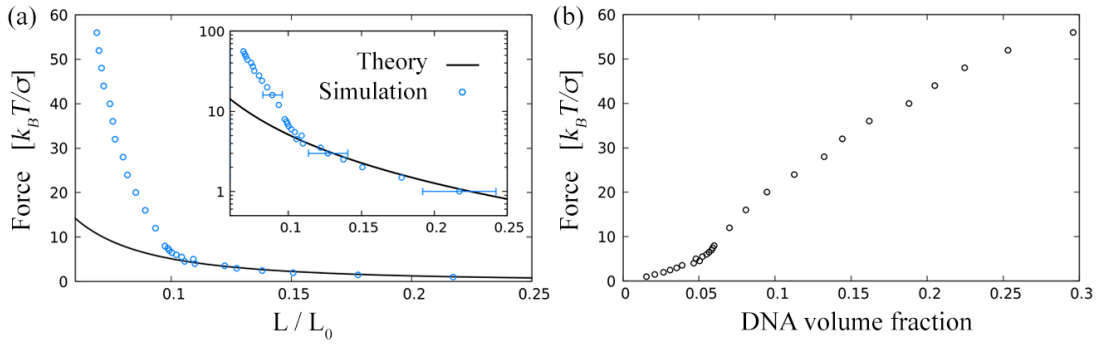

Figure 4.4 *(a) Force-extension curve for DNA in the absence of proteins, for the BD simulations (blue circles) and the entropic spring theory of Ref. [104] (full line). For $L/L_0 > 0.1$, the theory agrees with the model in that there is a strong dependence of the DNA extension on the compression force. For $L/L_0 < 0.1$, the theory deviates from the numerical estimate of the entropic force, confirming that, in the high compression regime, the assumption done in the theory of a linear chain of polymer blobs of diameter D breaks down. (b) Compression force as a function of the DNA volume fraction. The fact that the volume fraction reaches almost 0.3 for the tightest compression suggests that close packing and many body interactions are the most relevant contributions to the divergence of the numerical estimates from the theory in the high compression regime observed in plot (a).*

This result suggests that steric repulsion is an important factor to consider when estimating the entropic force resisting compression. At the same time, we note that the effect found here (a 10-fold increase) is an overestimate of the correction needed for a real bacterial chromosome. This is due to the scaled down dimensions of the DNA we use: the same value of L/L_0 corresponds to a much denser volume fraction of DNA in the simulations with respect to the experiments in Ref. [104].

4.3 Quantifying the effect of proteins

4.3.1 Molecular crowders greatly increase the entropic force exerted on the compression piston

To analyse the effect of bacterial proteins on the entropic elasticity of DNA, the simulations performed in the previous section were first repeated considering the presence of crowding proteins in the system. The osmotic pressure of proteins

in the cytosol can be estimated as Mk_BT/V, where M is the number of proteins, and V is the confinement volume. For the case of *E. coli* this corresponds to ~ 0.01 atm, which is larger than, or at least of the same order of, the pressures recorded in the experiment in Ref. [104].

Figure 4.5(a) shows how the presence of non-DNA-binding proteins affects the force-extension curves. We find that proteins make an important contribution, especially for moderate compression (relatively large values of L): in this regime, the force required for a given extension is orders of magnitude larger than in the case where the proteins are not included. More specifically, the force is approximately linear in the number of proteins M, as expected for an osmotic contribution (see Fig. 4.5, inset 1). Interestingly, for intermediate forces, we find that the extension recorded at a given force also depends linearly on M (Fig. 4.5, inset 2).

Figure 4.5(b) compares the numerical results for $M = 2000$ with the force resulting from an ideal gas of 2000 proteins (and no polymer, dashed line). Pressures from the theory are converted into forces along the cylinder axis via multiplication by the cross section of the cylinder. There is overall a reasonable agreement, which quantitatively confirms that the compression curves are dominated by the osmotic pressure of proteins. The agreement improves when correcting for the finite volume of proteins by using a pressure given by the van der Waals equation of state (solid line). The remaining discrepancy in the high compression regime may be due to the presence of the polymer, or to excluded volume effects not captured correctly by the van der Waals formula.

Next, the role of DNA-binding proteins is quantified by repeating the compression simulations considering 3% of the proteins to be DNA-binding. These tend to compact the DNA, and should lead to a change in the extensional elasticity. Mobile cross-links, or slip-links, have previously been shown to strongly modify the force-extension curves of polymers in stretching experiments [91]. For this reason, DNA-binding proteins were proposed in Ref. [104] as a possible explanation for the quantitative significant discrepancy between the experiments and the entropic spring theory regarding the polymer spring constant. Therefore one can ask to what extent the presence of DNA-associating proteins affects our results. Fig. 4.6 shows that these effects are minor in the simulations. This is consistent with our previous finding that osmotic forces from non-DNA-binding proteins are more important quantitatively than DNA entropic forces: DNA-

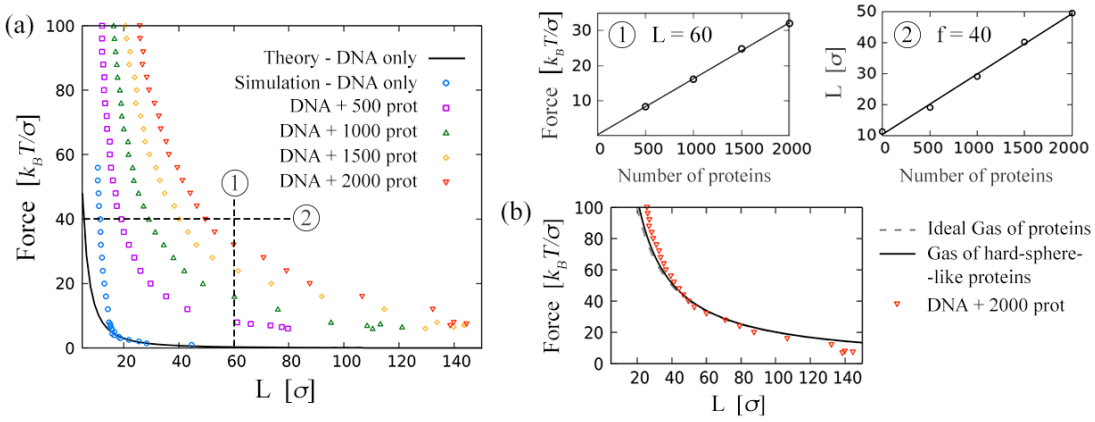

Figure 4.5 *(a) Quantification of the effect of crowding proteins on the numerical results for the compression force as a function of the DNA extension. The protein osmotic contribution leads to a large increase in the compression force, sometimes of several orders of magnitude. For a given value of L, the protein contribution to the compression force is linear in the number of proteins (inset 1) – fit:* $f = 0.0160M + 0.2933$. *For a given compression force, the DNA extension also increases linearly with the number of proteins (inset 2) – fit:* $f = 0.0195M + 10.4113$. *(b) Comparison between numerical results (orange inverted triangles) and predictions from the theory for an ideal gas of* 2000 *non-DNA-binding proteins (dashed line), and from a theory accounting for the volume of the protein through the van der Waals equation of state (solid line). The dashed and solid curves are nearly coincident with the numerical results.*

binding proteins only affect the polymer response, and hence do not have much bearing on the overall curve. One should stress here that the DNA-binding proteins we consider here form clusters, as could be the case for H-NS [17, 32] or DPS [141]; while the dynamic crosslinks invoked in Ref. [104] will in practice interact entropically (see [90]), their collective behaviour may be different.

In summary, we find that crowding proteins exert a significant osmotic pressure on the piston, and this is much larger than the force exerted by the DNA. Furthermore, any reduction in the force exerted by the DNA due to DNA-binding proteins is dwarfed by the contribution of the non-DNA-binding proteins. Our estimates suggest that even in the experimental situation the presence of non-DNA-binding proteins could significantly affect the force measured via the set-up used in Ref. [104]. Therefore it would be of interest to compare in more detail those experiments with *in vitro* compression experiments with different sized DNA molecules and different protein environments.

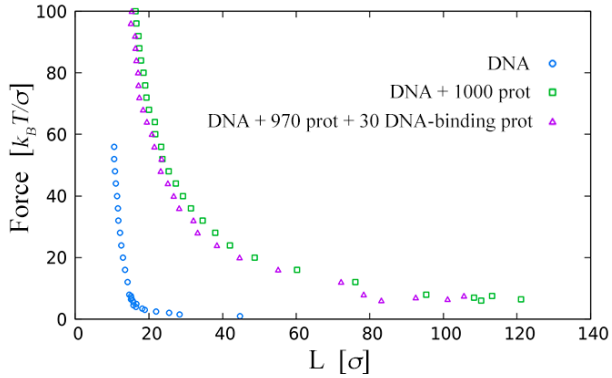

Figure 4.6 *Compression force as a function of the DNA extension for DNA-only simulations (blue circles), simulations in the presence of non-DNA-binding proteins (green squares) and of both types of proteins (orange triangles): 3% of the proteins are DNA-binding. The DNA-binding proteins lead to formation of DNA clusters, hence compacting the DNA, which yields a decrease in the compression force. The effect in the osmotic pressure due to DNA-binding proteins is, however, small.*

4.3.2 DNA expansion dynamics is hampered by molecular crowding and DNA-protein interactions

The dynamical behaviour of a DNA molecule (with or without proteins) is here examined by performing simulations where the DNA is first compressed under a strong force, and then let free to expand after the piston is removed. Figure 4.7 shows how the extension L increases in time following the piston removal. Note that the curves for a single realisation are very noisy, underscoring the highly stochastic nature of the expansion dynamics. Therefore, in Figure 4.7 each curve (or point) corresponds to an average over 10 independent runs.

In Section 2.2.3, I revised the scaling argument used to infer the time dependence of the polymer extension L, in the expansion regime. This simplified theory works in the adiabatic limit where the polymer goes through quasi-equilibrium states, i.e. for slow dynamics. The derived power-law scaling relations thus provide a good scheme for interpretation of the present Section's results, as we focus on the dynamics of the polymer whilst it relaxes after the initial abrupt expansion. In Appendix C, I further present results from simulations performed for a longer DNA polymer (with $N = 2000$ beads) which corroborate the assumption that, after the initial abrupt expansion, the polymer extension along the cylindrical pore increases with time in a power-law fashion.

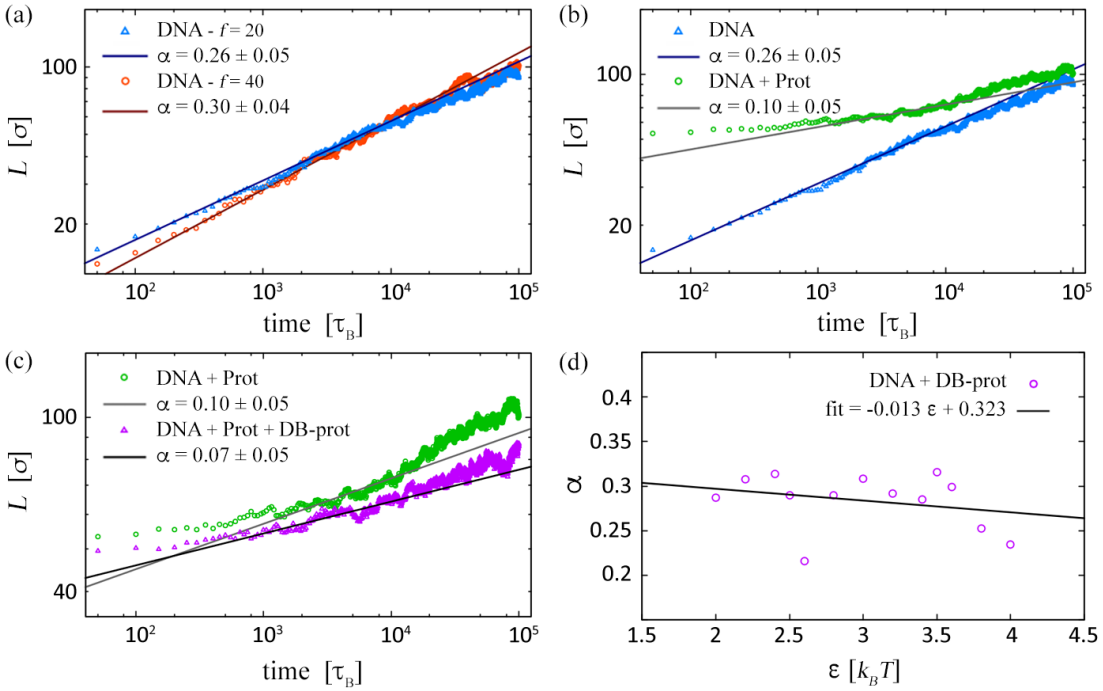

Figure 4.7 *Measured dynamical exponents during DNA expansion (a) for DNA-only simulations, for different initial compression forces: $f = 20\ \epsilon/\sigma$ (blue triangles) and $f = 40\ \epsilon/\sigma$ (orange circles); (b) for simulations in the presence of* 1000 *crowding proteins (green circles) – protein crowding leads to a decrease in the exponent; and (c) for simulations in the presence of* 970 *crowding and* 30 *DNA-binding proteins (magenta triangles) – the exponent decreases in the presence of DNA-binding proteins, suggesting that the formation of protein-induced DNA clusters slows down the DNA expansion. (d) Measured exponents for simulations with DNA and* 100 *DNA-binding proteins, but no crowding proteins, as a function of the DNA-protein affinity. The exponents show a weak tendency to decrease with increasing affinity, supporting the observation that DNA clustering slows down DNA expansion. All simulations started from an initial DNA configuration obtained for a compression force of* 20*, except for the case in the top-left plot. The ranges of times, for which the expansion curves were fitted, were chosen to take into account the intermediate expansion regime ($10^3\tau_B$ – $10^4\tau_B$ or* 0.01 *ms –* 0.1 *ms), which corresponds to a power law behaviour for all forces we considered. For the curves plotted in (a-c) the error bars are comparable to the size of the data points. In plot (d), where each data point corresponds to a single fit, the error on the fit is not used since this is underestimated.*

Therefore, for the DNA-only case, one should expect a scaling behaviour for L with $t^{4/13}$. The simulation results typically show a similar, but on average slightly smaller, exponent: the values are also consistent with those found numerically in Ref. [61]. Expansion simulations were performed starting from two different initial compression forces: $f = 20\ \epsilon/\sigma$ and $f = 40\ \epsilon/\sigma$. Results in Figure 4.7(a) show a slightly different apparent exponent for the different initial conditions: the

more compressed the initial DNA configuration, the bigger the apparent value of the exponent. Note however that this difference between the exponents is small, or even not significant within the exponents' error bars. Therefore, this may either point to some dependence on the initial condition, or to a variability of the exponent (which is apparent from Fig. 4.7(d) and further discussed below).

The effect of proteins is addressed by repeating the expansion simulations in the presence of 1000 proteins. Fig. 4.7(b) compares the results for the DNA-only and the DNA plus crowding proteins simulations. The expansion dynamics is much slower (i.e., the apparent exponent is actually smaller) in the presence of non-DNA-binding proteins: this is because the proteins create a crowded environment which hampers DNA unfolding. At the same time, the DNA extension is larger at $t = 0$, for the same initial force, because of the osmotic pressure of the proteins which opposes the compression force from the piston. As a result the value of $L(t)$ is always larger when proteins are present in the simulations. For late times, the proteins diffuse away from the DNA and become dilute, so the effect of crowding diminishes. Hence for large t the apparent exponents for the curves with and without proteins are expected to become approximately equal.

DNA-binding proteins lead to even slower progress in the DNA expansion (Fig. 4.7(c)), although the change in the exponent α, with respect to the case with just crowding proteins, is small or even not significant within the exponents' error bars. Therefore, to characterise this effect more in detail, we performed simulations with DNA in the presence of *only* 100 DNA-binding proteins (Fig. 4.7(d)): the exponents found in this case should then be compared with those for DNA-only simulations (Fig. 4.7(a)). These results confirm that when proteins bind to the DNA they may reduce the apparent exponent; however the effect is subtle, and there is a large stochastic element in the dynamical curves (as each of the values in Fig. 4.7(d) is computed by averaging 10 different expansion runs).

4.3.3 "Popping-off" dynamics with DNA-binding proteins

In Figure 4.7(c-d), we analysed the effect of DNA-binding proteins on the dynamics of DNA expansion [2]. By examining the trajectories of the system for different DNA-protein affinities, some further interesting phenomena in the kinetics of the system are observed.

This is illustrated in Figure 4.8. If the DNA-protein interaction is weak ($\epsilon = 2k_BT$), proteins only transiently bind to DNA. Since increased concentration favours the bound state (the entropy loss upon binding is smaller), more proteins bind upon compression. However these detach immediately after the piston is released (see Fig. 4.8, top row, and blue line in side panel). As a result, the DNA responds elastically as a protein-free polymer – this is consistent with the finding that the apparent exponent measured for a low DNA-protein interaction strength ϵ, in the simulations with only DNA-binding proteins, is comparable with the exponent measured for the DNA-only simulations (see Fig. 4.7).

Some interesting dynamics occur if one chooses a larger DNA-protein interaction strength, so as to promote more long-lived binding (see Fig. 4.8, middle row). In this situation, the interaction strength is such that it favours long-lived binding under compression. In this case, essentially all proteins are bound at all times, and they locally compact the DNA into a toroidal structure, resembling that of DNA within bacteriophages [3]. When the piston is removed, the translational entropy of the proteins in the unbound state increases dramatically – as they could now occupy any region of the cylindrical domain. However the bound state is still metastable, and it takes a relatively long time for the proteins to detach. Over time, proteins "pop-off", typically one-by-one, from the collapsed DNA. This can be seen in Figure 4.8, as the overall pair energy of the polymer-and-protein system (which is approximately proportional to the number of bound proteins) decreases in magnitude linearly with time. Therefore, the "popping-off" time should increase linearly with number of DNA-binding proteins in the system, so could easily be observable in experiments with bacterial DNA (recall there are an estimated 3×10^4 DNA-binding proteins in the bacterial nucleoid). Intriguingly,

[2]Note that these results correspond to proteins which bind non-specifically to DNA. It would be of interest to ask what is the effect of *specific* binding, which is known to promote the formation of clusters with a well-defined size [14].

[3]In a real bacterial DNA, proteins like H-NS would not bind the whole genome so the 3D structure would likely be different upon compression. However, we expect similar dynamics to occur there as well.

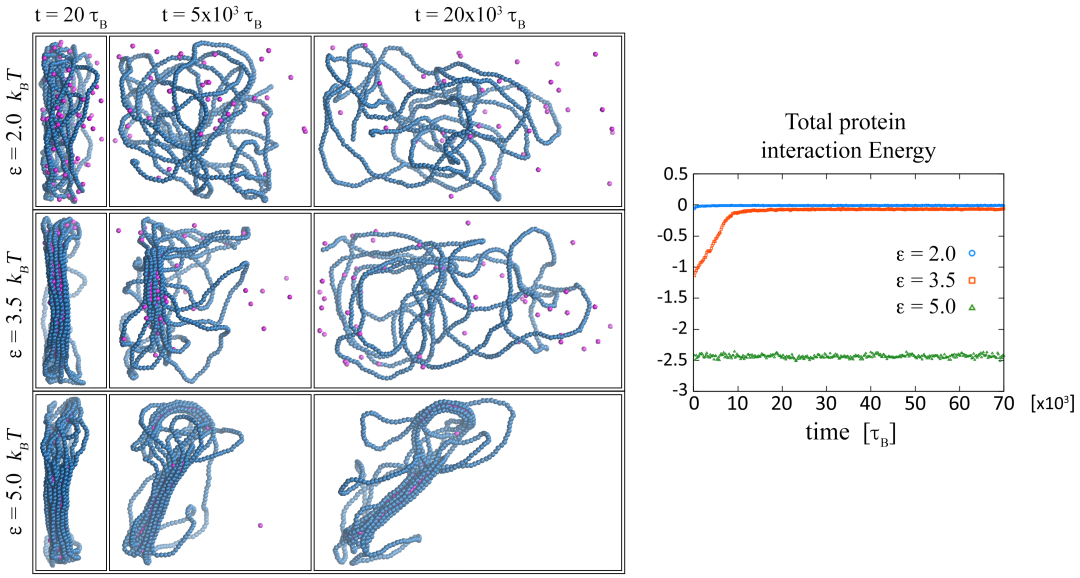

Figure 4.8 *Protein popping-off during DNA free expansion, for simulations with DNA and* 100 *DNA-binding proteins. The right plot shows the overall pair energy of the system as a function of time. This quantity is negative and approximately proportional to the number of proteins bound to the DNA at a given time. For low DNA-protein interaction strength ($\epsilon = 2.0$), the proteins do not stick to the DNA, as confirmed in the energy-time plot: the pair energy remains zero over time. For large DNA-protein interaction strength ($\epsilon = 5.0$), the proteins remain permanently bound to the DNA during expansion, and the pair energy remains approximately constant over time. For an intermediate interaction strength ($\epsilon = 3.5$), after removing the piston some proteins remain bound to the DNA, but eventually pop off. The pair energy first decreases in magnitude linearly in time and then tends to zero asymptotically.*

while the protein kinetics are completely different, the popping-off leaves little detectable signature in the apparent exponents recorded in Figure 4.7. Finally, the popping-off requires tuning the interaction energy, because if this becomes too large, proteins bind to the DNA permanently (at least within the simulation time considered here), and the popping-off kinetics can no longer be observed – the energy now does not appreciably depend on time (see Fig. 4.8, side panel, green curve).

4.4 Remarks

In summary, in this Chapter I have presented Brownian dynamics simulations of the compression and expansion dynamics of a DNA molecule, modelled as a self-avoiding linear polymer, interacting with an ensemble of crowding and DNA-

binding proteins. Previous work [17, 35, 56, 69, 125] had clearly demonstrated the importance of proteins on the thermodynamic conformations of bacterial DNA: non-DNA-binding proteins create macromolecular crowding which can promote global collapse of the bacterial chromosomes, while DNA-binding proteins such as H-NS provide local compaction. The novelty of the presented work is that, by addressing a set-up which is directly relevant to single molecule experiments probing the entropic elasticity of bacterial DNA, one can establish that the inclusion of proteins further leads to important effects on the force-extension curves recorded upon compression, and also on the expansion dynamics of the polymer.

First, I presented DNA-only simulations, where there are no proteins, as a reference case. I have shown that for weak to intermediate compression the numeric results confirm the entropic spring theory of Ref. [104], which predicts an abrupt decrease in elongation with compression. For large compression ($L/L_0 < 0.1$ in the simulations), however, excluded volume interactions create a strong deviation from the theory.

Second, the presented results reveal that the osmotic pressure of non-DNA-binding proteins can dwarf the entropic pressure of the spring-like polymer during compression. In the simulations, a polymer of $N = 1000$ "beads" was considered in the presence of up to $M = 2000$ non-DNA-binding protein "beads". Although these numbers are comparable, the effect of proteins can be orders of magnitude larger in the compression curves. With respect to these thermodynamic curves, DNA-binding proteins affect the DNA elasticity, hence only have a minor effect overall.

It is interesting to ask whether osmotic pressure from unbound proteins can account for the high force (~ 100 pN) observed experimentally in the compression curves in Ref. [104]. As mentioned by the authors, while the curve overall can be fitted with the entropic spring theory, the numerical value which is expected of a self-avoiding polymer would be significantly smaller than observed – the work presented in this Chapter suggests that even a small fraction of proteins could lead to an osmotic force which might account for this.

Third, the impact of proteins (again, non-DNA-binding and DNA-binding), on the DNA expansion dynamics was explored. The DNA-only simulations lead to a scaling behaviour for the extension L as a function of time which is consistent with previous work in the literature [61]. However, the crowding introduced

by non-DNA-binding proteins leads to a much slower dynamics, and to a much decreased apparent exponent. An interesting observation is that, when simulating DNA with DNA-binding-proteins only, tuning of the DNA-protein affinity leads to a popping-off kinetics during DNA expansion, where proteins are metastably bound to the DNA and detach one-by-one after the volume at their disposal increases.

While the results presented in this Chapter focus on the case of proteins which are not charged, and have the same size as DNA beads (a choice made for simplicity), we have also performed simulations taking into account the fact that proteins may have a non-negligible charge, and that the size of a typical bacterial protein is larger, and about twice the thickness of B-DNA [106]. The results are presented and discussed in Appendix A for charged particle, and in Appendix B for a protein size of 2σ. In all cases, the trends for the different results are qualitatively identical, and confirm the above conclusions. There are quantitative differences for the size simulations when the volume fraction is high: this is expected as, under those situations, it is the volume fraction, rather than number density, which determines the compression pressure and force. It is interesting that the effective exponent for DNA expansion slightly decreases for larger proteins, due to the increased crowding.

Overall, the proposed model and subsequent results provide a generic framework within which to analyse experiments such as those in Ref. [104]. They also provide further testable predictions for future experiments, e.g., probing the kinetic dynamics during the expansion of bacterial DNA *in vitro*: since the protein "popping-off" time is expected to increase linearly with the number of DNA-binding proteins, this could be easily observable and measured for bacterial cells (which contain $\sim 3 \times 10^4$ of DNA-binding proteins) and subsequently tested by manipulating the number of proteins in the system.

5 3-D structural organisation of human chromosomes

The 3D spatial structure of chromatin and chromosomes *in vivo* during interphase is intimately linked to gene regulation and cellular integrity [1, 23, 142], as discussed in Chapter 1. Chromatin structure has been shown to change during development as cells differentiate, in ageing or senescence [30], and in disease [81]. Recent years have seen major developments in a number of techniques to investigate the 3D chromosomal conformation in different organisms and cell types during interphase [37, 78, 111, 124] and mitosis [96]. The most widely employed technique to date is "HiC" – a high-throughput and genome-wide version of "chromosome conformation capture" – which uses cross-linking, restriction enzymes, and ligation with biotin-labelling, followed by high-throughput sequencing, to build maps quantifying the probability of interaction between different genomic loci within a population of cells [37, 78, 111, 121, 124]. These maps, constructed for different organisms and cell types [37], naturally lend themselves to comparison with those predicted by "bottom-up" computational models based on polymer physics principles [14, 45, 97].

Notwithstanding the wealth of new experimental data on 3D genome organisation, the mechanisms which drive structure formation are still not fully understood. Two main classes of biophysical models are currently popular in the field: the "transcription factor" (TF) model [13, 14, 17, 98] (also known as the "strings-and-binders" model [7]); and the "loop extrusion" (LE) [2, 45, 120] model.

The first (Fig. 5.1(a)) postulates that multivalent chromatin-binding proteins mediate chromatin-chromatin interactions, creating loops and driving 3D folding (see Sec. 2.3.2 for more details). Examples of proteins that play key roles within this framework are transcription factors associated with active chromatin (eu-

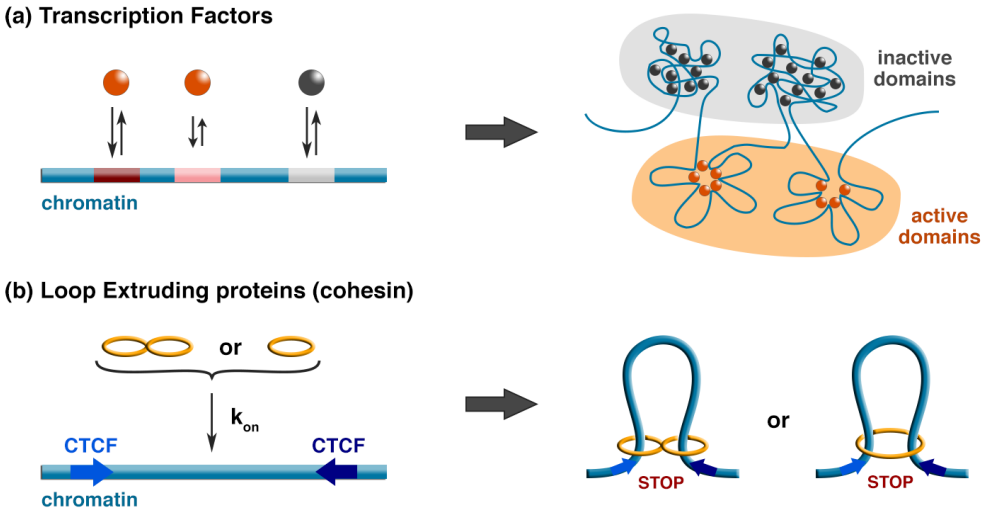

Figure 5.1 *(a) Transcription Factors bind to their genome cognate sites, and naturally organise the chromatin into domains through the "bridging-induced attraction". Domains of different types phase separate so that active regions are more likely to interact with other active regions, and inactive with other inactive. (b) Loop extruding proteins bind to the genome and move along the chromatin fibre, effectively extruding chromatin loops. It remains unclear whether cohesin acts as a single extruding ring [45] or as a pair of rings [120], as well as whether it moves actively or "diffuses" along the genome.*

chromatin) [27], polycomb group proteins associated with facultatively repressed regions (facultative heterochromatin) [4], and HP1 or related proteins associated with constitutively repressed regions (constitutive heterochromatin) [74] [1]. This model naturally explains the large-scale (micro)phase separation[2] of the genome into active and inactive domains [14] (where active regions are more likely to interact with other active regions, and inactive with other inactive [78]) and the formation of nuclear bodies [16], both driven by the "bridging-induced attraction" mechanism [17] (see Sec. 2.3.2). The latter is consistent with recent works which have revealed that heterochromatin forms liquid-like phase-separated regions [74, 131], and that many transcription factors form discrete foci incorporating their DNA binding sites [26].

The second model (Fig. 5.1(b)), which has recently come to the fore, posits

[1] Constitutive heterochromatin corresponds to repressed chromatin regions that remain condensed throughout the cell cycle. Facultative heterochromatin corresponds to repressed regions where active chromatin is condensed into heterochromatin, and, thus, has the potential for gene expression, being condensed or decondensed depending on the cell stage.

[2] Here microphase separation refers to phase separation into clusters which only coarsen up to a certain size.

that the SMC[3] complex cohesin and the CCCTC-binding factor (CTCF) are the master organisers of the genome, suggesting that cohesin acts as a loop extruding factor [122] which actively creates expanding loops, but halts when it meets a bound CTCF. This model can account for the striking bias in favour of convergent CTCF loops [111] and it can also rationalise the "topologically-associated-domain" (TAD) patterns observed in HiC maps [45]. However, no motor activity has yet been found in experiments probing the motion of DNA-bound cohesin *in vitro* [33, 65, 130] – in contrast, the structurally related condensin complex has been shown to be able to move unidirectionally along DNA [133]. The convergent loop bias can also be explained by a model of diffusive loop extrusion (dLE) where cohesin slides diffusively along the chromatin rather than actively moving unidirectionally [15]. Another possibility is that the diffusive motion is enhanced by ATP consumption resulting in an active bidirectional motion.

The TF and (d)LE models each explain different aspects of genome organisation, suggesting that they may be complementary, and both important for chromosome organisation. While the TF model describes a functional level of genome organisation, intimately linked to the local transcriptional activity and chromatin state, the LE model describes a level of organisation independent of these. If the reality is a combination of both models, one would expect that experiments disrupting either transcription factors or cohesin binding should give rise to distinct changes in chromosomal architecture. Overall this is the case, although the observations are not at all straightforward. Very high resolution conformation studies of the globin loci using Capture C [13, 51] revealed completely different conformations in erythroid cells, where these genes are very active, and stem cells, where they are inactive – i.e. changes in protein binding sites result in changes in conformation. Likewise, cohesin or CTCF knock-outs result in the disruption of the observed loops and loop-domains [71, 99, 110, 122], but appear to leave the underlying chromatin states mostly unchanged [99].

In this chapter I will single out the different roles of transcription factors and loop extruders by comparing the TF and (d)LE models in terms of their ability to predict chromosome organisation.

I will, therefore, focus on a 30 Mbp section of human chromosome 7, and show

[3]Structural Maintenance of Chromosomes (SMC) proteins consist of a family of proteins involved on the higher-order chromosome organisation and dynamics.

that neither the TF nor the LE model can, by itself, give a satisfactory account of the observed folding of the entire chromosome segment. The (d)LE model accurately predicts the chromatin domain pattern locally, but fails to capture larger-scale interactions. On the contrary, the TF model poorly predicts the fine detail of local interactions, but captures long-range contacts more faithfully.

The results in Section 5.2 show that, instead, a combination of the TF and (d)LE models can reproduce many of the HiC features, suggesting that TFs and cohesin (or other LE factors) indeed have complementary roles in genome organisation. I will show that LEs are required to fold regions of "inert chromatin", where there is a void of active or repressive histone modifications and, hence, transcriptional activity is sparse. On the other hand, the simulations suggest that TFs are sufficient to organise active/inactive domains, where cohesin-mediated loops play a more minor role.

I will then investigate the intriguing fact that a straightforward superposition of the standard TF and (d)LE models still leaves some key qualitative discrepancies between simulated and HiC interaction maps, i.e., the simulations show too high a signal for medium to long range interactions. For that, I will consider a more complex version of the model where TF binding is modified by a non-equilibrium mechanism, where TFs constantly switch between an "on" (binding) and an "off" (non-binding) state [16]. This switching-TF (sTF) model encodes a dynamic level of control on TFs and it might represent post-translational modifications of the proteins affecting their binding affinity to chromatin [94], or protein turnover through active degradation and replacement. In Section 5.3, I will show that the combined sTF and LE model actually gives better predictions of chromosome interactions, and that it is also consistent with single-molecule microscopy experiments on the dynamics of active/inactive chromatin domains, chromatin loops and protein clusters [24, 49, 134].

In Section 5.4, I will look deeper into the mechanism of chromatin loop extrusion, and compare the performance of models for active and diffusive loop extrusion. Since 3-dimensional MD simulations for diffusive extrusion require unfeasibly long times or coarser polymer resolutions, I will instead consider a 1-dimensional lattice dLE model, where diffusive extruders obey the same extrusion rules as in the 3-D simulations. I will then compare the results for active 3-D and diffusive 1-D extrusion, and show that diffusive extrusion can account for the formation of chromatin domains and CTCF loops.

Finally, in Section 5.5, I will explore the potentialities of the combined sTF and LE model, by simulating recent protein knock-out experiments [99, 122]. The results show that the model can actually correctly predict the main experimental observations.

5.1 Modelling chromatin folding by transcription factors and cohesin

5.1.1 The system

In this Chapter, the first 30 Mbp of chromosome 7 of the human lymphoblastoid cell line GM12878 are modelled, for which high-resolution HiC data are available [111]. This region includes large gene deserts (regions of "inert chromatin", where transcriptional activity is sparse and which is void of active or repressive histone modifications), as well as facultative and constitutive heterochromatin, and active regions (see Fig. 5.2). Therefore the model focuses on a single cell type during interphase, where the pattern of the different chromatin states (i.e. the epigenetic landscape) are already established, and, thus, it does not take into account the dynamic epigenetic changes that occur during cell differentiation or right after cell division.

The 3-D structural properties of the chromosome segment are probed by simulating the folding of the segment in the presence of transcription factors and/or chromatin extrusion proteins (cohesin).

Note that, since we are modelling just a segment of the human chromosome 7, strong confinement could lead to unrealistic long-range chromatin interactions. Therefore the system is modelled in the dilute regime so as to avoid non-physical confinement effects.

5.1.1.1 Chromatin Polymer

The chromosome segment is modelled, at the resolution of the chromatin fibre, as a linear self-avoiding polymer composed of spherical beads connected by spring bonds, according to the details described in Chapter 3. Each chromatin bead, with diameter $\sigma = 30$ nm, corresponds to 3 kbp of DNA, hence the polymer is

10,000 beads long. The persistence length is set to $l_p = 3\sigma = 90$ nm, which is in the range of measured values for chromatin, $30 - 250$ nm [73].

Polymer beads are "coloured" according to their underlying chromatin state based on histone modifications. The latter correspond to post-translational modifications to histone proteins that change their interaction with other proteins and DNA. Histone proteins have tails that extend out of the nucleosome (see Fig. 5.2(a)). These are composed of residues that can be covalently modified by methylation, acetylation and phosphorylation, among others. Different modifications are associated with different transcriptional states of DNA loci.

In this study different genomic regions are considered, corresponding to transcriptionally active enhancers and promoters, transcribed genes, constitutively and facultatively repressed genes. These regions are particularly enriched, respectively, in the histone modifications H3K4me1, H3K4me3, H3K36me3, H3K9me3 and H3K27me3. The genomic location of the different modification marks in GM12878 cells is obtained by using ChIP-seq[4] data obtained from the ENCODE project [40] (http://www.encodeproject.org [127]).

Polymer beads were then annotated as active (enhancers or promoters), transcribed, heterochromatic or polycomb depending on their overlap with the called ChIP-seq peaks. It is possible that some beads overlap multiple peaks, so these can have multiple annotations. For instance, there can be beads that correspond to binding sites of active and repressive proteins. This does not mean that the different proteins bind to the same genomic sequence, but that the polymer bead in the model corresponds to a genomic locus with sequences bearing different modification marks. Regions of the genome that are not enriched in any particular mark are defined as "inert", and left unmarked. Inert chromatin is AT-rich and gene-poor, so that it bears some of the signatures of heterochromatin, though it is not characterised by an enrichment of either the H3K27me3 or H3K9me3 histone modifications. In this way the modelled chromatin segment becomes a co-polymer whose segments interact with freely diffusing bridge-forming protein complexes – the TFs.

[4]Chromatin Immunoprecipitation Sequencing (ChIP-seq) is a technique used to identify the DNA binding sites of targeted proteins. By using specific protein antibodies, DNA loci with bound proteins can be isolated and later sequenced. This technique then yields a profile of the protein binding probabilities along the genome. The genomic location of histone modifications can, therefore, be determined by using specific proteins known to bind to the different modifications.

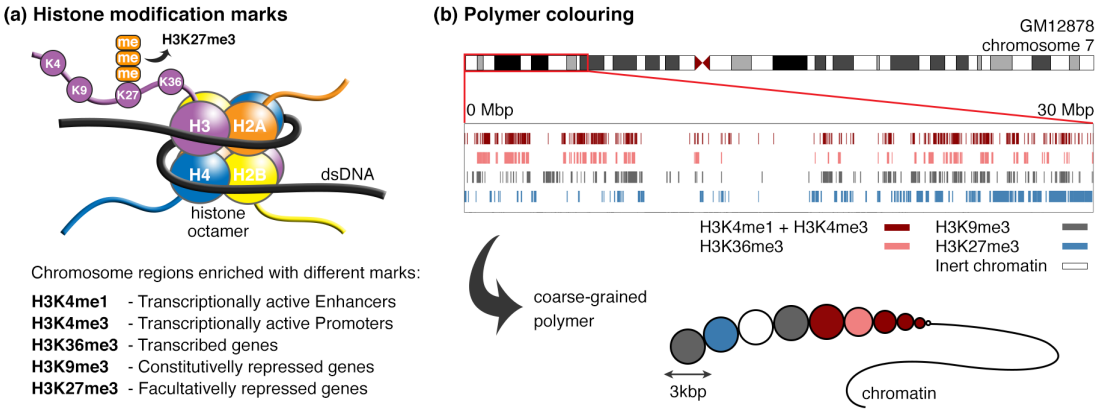

Figure 5.2 *(a) Different chromatin regions (active/repressed genes) are enriched in different histone modifications, so that these serve as marks of the underlying chromatin state. (b) In the present model, each polymer bead (3 kbp) is "coloured" according to the abundance, in the respective chromatin locus, of the modifications: H3K4me1, H3K4me3, H3K36me3, H3K9me3, and H3K27me3. Histone modification ChIP-seq data for a human lymphoblastoid cell line GM12878 are obtained from the ENCODE project [40].*

5.1.1.2 Transcription Factors

Transcription factors are modelled, for simplicity, as spheres of diameter $\sigma_{\text{prot}} = \sigma$. The interaction between TFs is purely repulsive and described by the WCA potential in Eq. (3.5). Three species of TFs are considered (see Fig. 5.3(a)): (i) euchromatin-binding, which bind to promoters, enhancers and transcribed genes; (ii) heterochromatin-binding, such as HP1; and (iii) polycomb-group proteins, such as PRC[5]. A total number of $M \sim 1400$ proteins is considered [(i)∼250, (ii)∼300, (iii)∼850], meaning that for a cubic simulation box of size $L = 220\sigma$ the volume fraction occupied by particles is ∼0.05%. Bearing in mind that in typical cellular conditions the nuclear volume fraction occupied by nuclear organelles is around 15%, the modelled system is in the dilute regime. This choice was purposely made, as mentioned above, so as to avoid non-physical confinement effects (see Sec. 5.1.1). The number of proteins of each type was chosen according to the relative proportion of euchromatin, heterochromatin and polycomb-binding polymer beads.

TFs bind to their cognate chromatin beads through an attractive interaction set by the LJ potential in Eq. (3.7), with $r_{\text{thr}} = 1.8\sigma$. Euchromatin-binding

[5]PRC stands for Polycomb Repressive Complex, a protein complex binding to chromatin regions enriched with the H3K27me3 mark.

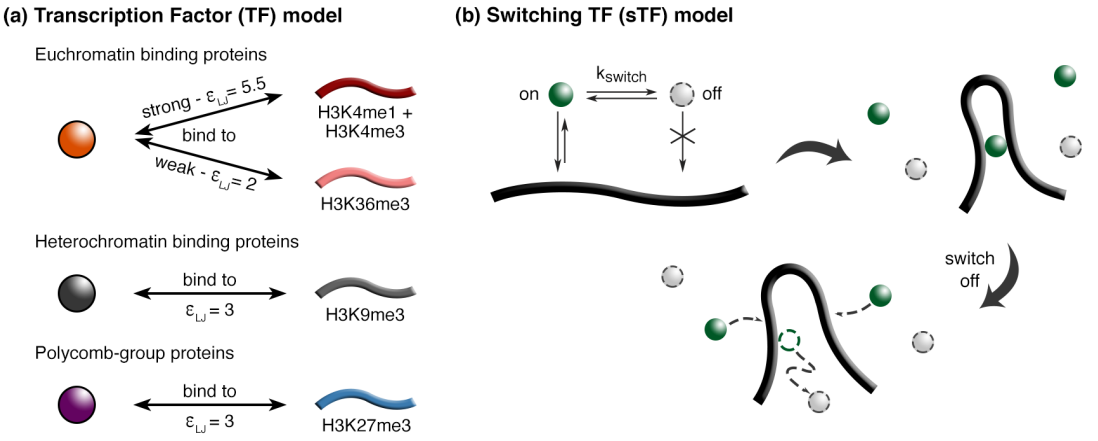

Figure 5.3 *(a) In the TF model, multivalent chromatin binding proteins organise the chromatin by forming molecular bridges between their different binding sites. Here we consider three species of TFs, whose cognate chromatin binding sites are inferred from histone modifications (see Fig. 5.2). (b) The TF model can be enhanced by implementing TF switching: TFs switch between an "on" (binding) and an "off" (non-binding) state at rate k_{switch}. This leads to the formation of dynamic, rather than static, protein clusters.*

TFs interact strongly, $\epsilon = 5.5k_BT$, with promoters and enhancers and weakly, $\epsilon = 2k_BT$, with transcribed genes. Heterochromatin-binding TFs interact with moderate strength, $\epsilon = 3k_BT$ with heterochromatin and interact weakly, $\epsilon = 2k_BT$, with polycomb regions. Finally, polycomb-group proteins interact with moderate strength, $\epsilon = 3k_BT$, with polycomb regions and interact weakly, $\epsilon = 2k_BT$, with heterochromatin. Thus, during the course of a simulation, TFs can bind and form bridges between chromatin beads bearing the associated histone modification marks.

As mentioned earlier in this Chapter, in this study we will also look at a TF model where post-translational modifications (PTM) of TFs are considered, by allowing TFs to dynamically switch between an "on" and "off" state. While in the "on" state, TFs bind chromatin through a Lennard-Jones potential as in Eq. (3.7). When in the "off" state the interaction reverts to the non-attractive WCA potential (Eq. (3.5)). Importantly, the transition between these two states is controlled by an additional parameter (the switching rate k_{switch}) that is not related to the equilibrium transcription factor unbinding time, $\sim \tau_B \exp\{\epsilon/k_BT\}$. Thus, the binding affinity ϵ and the switching time k_{switch}^{-1} can be tuned separately (e.g. we can have $\epsilon \gg k_BT$ while $k_{\text{switch}}^{-1} < \tau$). Since switching is independent of the 3-D conformation (and whether the TF is bound or not), it drives the

system away from equilibrium, allowing behaviour that cannot be reproduced by a purely thermodynamic model [16]).

Note that the current coarse-grained model, where DNA is viewed as a polymer at the level of compaction of the chromatin fibre, does not consider explicitly the transcription process where TFs unwind the DNA and process along a single strand while transcribing the gene coded in that DNA sequence. Instead, TFs are here modelled as diffusing beads that, when in the "on" state, simply bind/unbind thermodynamically once in contact with their cognate chromatin sites. Under this simplification, made for the purpose of considering a coarse-grained chromatin polymer, the model still retains the key chromatin-TF interaction properties: TFs diffuse through the nuclear medium until they find a target chromatin site, where they transiently bind, in line with experimental studies probing the dynamics of transcription factors [46]; and euchromatin-binding TFs end up binding and bridging promoters and enhancers, stabilising loops, in agreement with experimental observations from HiC data [111].

5.1.1.3 Loop Extruding Proteins

Regarding chromatin loop extruding factors, two possible mechanisms are compared in this study: active loop extrusion, where the ends of the LE bond (LE "heads") move apart unidirectionally along the polymer at a specified speed; or diffusive extrusion (dLE), where the two LE heads move in either direction along the polymer with equal probability. For the active case, the extrusion speed is a parameter which can be chosen to give the best prediction of the experimental data. For diffusing LEs, if one chooses realistic parameters for diffusion, it takes much longer to generate chromatin loops of the same size - simulations would need to run for infeasibly long times. Recent work on chromatin loop extrusion mechanisms [15] (where much smaller chromatin segments are modelled) shows that a simple 1-D model, where LE heads diffuse along a lattice, can accurately capture the behaviour of more detailed 3-D simulations with explicit diffusing slip-links. Therefore, in this work, active extrusion will be studied using 3-D MD simulations, whereas diffusive extrusion will be studied using 1-D lattice simulations. Both studies can then be compared by performing a "fair" analysis, i.e. a 1-D-like analysis, of the 3-D simulation results (see Sec. 5.1.3.1). Here I shall start by considering active extrusion for the 3-D MD simulations (see below

for details for 1-D diffusive extrusion).

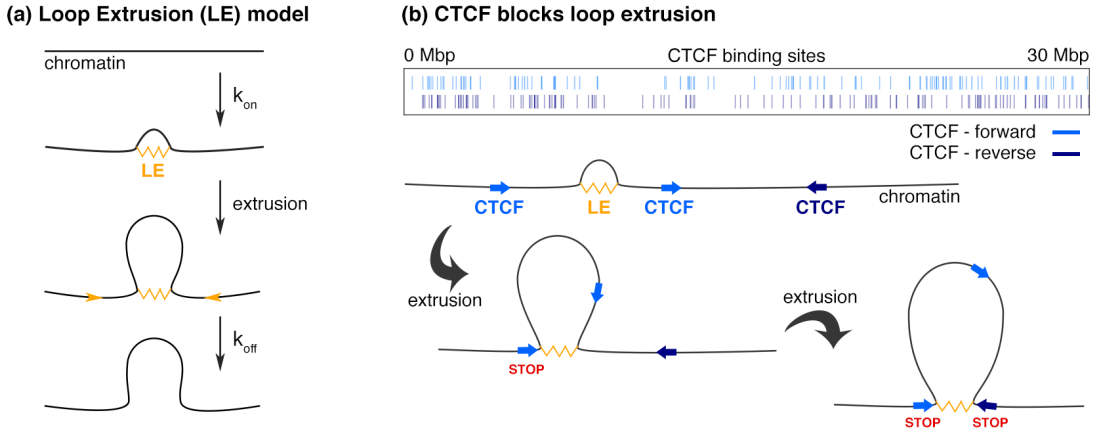

Figure 5.4 *(a) In the LE model, extruding factors (e.g. representing the cohesin complex) bind and unbind chromatin with rates k_{on} and k_{off} respectively. LEs are modelled as harmonic springs joining two polymer beads. Binding/unbinding is simulated by creating/deleting LE bonds. LEs also unbind if the bond becomes longer than $r_{ij} > 4\sigma$. While bound, LEs move along the chromatin, extruding a loop. (b) LEs stop moving if they reach a CTCF site where the binding motif is orientated towards its direction of motion, or if they encounter another extruder. ChIP-seq data for CTCF binding in GM12878 cells are obtained from ENCODE [40].*

Loop extruding factors are modelled in the 3-D MD simulations, for simplicity, as additional transient harmonic bonds that bring together two chromatin beads (see Fig. 5.4(a)). The bond is described by the harmonic potential in Eq. (3.3) with $r_0 = \sigma$ and $K_{\mathrm{LE}} = 6k_BT$. LE binding and unbinding is modelled by creating and deleting LE bonds. Bonds are initially created between i and $i+2$ chromatin beads, and a loop is extruded by "moving" the ends of the bond along the polymer, i.e., by deleting the bond and creating a new one that binds the next pair of chromatin beads.

An extruder comes to a halt either if it collides with another extruder or if it reaches a correctly oriented CTCF site (see Fig. 5.4(b)). This models the fact that cohesin interacts with the CCCTC-binding factor (CTCF) in an orientation dependent manner, favouring chromatin loops anchored by CTCFs with convergent motifs [45]. Since the CTCF consensus binding motif is non-palindromic, it has an orientation on the DNA. A pair of CTCF sites can therefore have one of four arrangements: the two motifs could point towards each other, away from each other, or both point in the same direction. Extruders are modelled as sensitive to this motif: if a bond end reaches a CTCF oriented in the opposite

direction to its motion, then it stops, otherwise it keeps on extruding. Therefore extruders end up bringing together CTCFs with convergent motifs, in line with experimental observations [111, 120].

ChIP-seq data for CTCF binding in GM12878 cells (Fig. 5.4(b)) were obtained from the ENCODE project [40] (http://www.encodeproject.org [127]). To account for the fact that there is a finite probability for CTCF binding at the identified sites, a binding probability was assigned to each site according to the called peak height. To model cell-to-cell variability in CTCF occupancy, in each repeat simulation a subset of the CTCF sites is populated based on these probabilities. The subset of site positions was then overlayed onto the 3 kbp/bead polymer. If more than one CTCF was present in a given bead, we assigned the orientation appearing most often.

An extrusion step occurs every 250τ (see Sec. 5.1.2 for comparison with real-time), meaning that the base line extrusion rate is 2 beads/250τ. The number of extruders bound to the polymer is kept at a constant value of 200: a new LE binds every time one unbinds. Extruders bind at randomly chosen positions along the simulated chromatin section and unbind with a rate $k_{\text{off}} = 0.0167\tau^{-1}$. In the case of an extruder which brings together a pair CTCF sites with convergent motifs, k_{off} is reduced by a factor of 10: $k_{\text{off, conv CTCFs}} = 0.00167\tau^{-1}$. The extrusion parameters were tuned so as to yield the formation of TADs as observed in the HiC interaction maps.

A popular candidate for the loop extruding factor is the cohesin complex [45, 120], and it is thought that a cohesin ring, or a connected pair of cohesin rings in a "hand-cuff" conformation, encircle two DNA segments, bringing them together. As such, the complex would have a finite volume and interact sterically with the surrounding chromatin and proteins. Since within our simplistic model, extruders do not occupy a finite volume (they are just transient bonds), we explicitly account for their steric hindrance by imposing a threshold on the maximum length between two beads connected by a LE bond. In practice, if the bond becomes longer than $r_{ij} > 4\sigma$, it is explicitly removed. This allows us to partially account for the inability of cohesin to move through a dense cluster of chromatin-bound to proteins, and constitutes a weak coupling between LEs and the 3-D conformation of the polymer and proteins. Note that previous work on extrusion [45, 100, 120] does not include this coupling.

For the 1-dimensional simulations of diffusive extrusion, chromatin is modelled

as a 1-D lattice with $N = 10,000$ lattice sites corresponding to the 10,000 simulated chromatin beads. The CTCF binding sites and motifs sequences used for the 3-D model are also used here, therefore repeat simulations are run for the same stochastically chosen subsets of CTCF sites. Diffusing LEs are modelled as two heads that move independently, each occupying a lattice site. Every simulation step, each LE head moves to either neighbouring site with equal probability. Like in the 3-D model, LE heads cannot go past each other, or go past a CTCF oriented oppositely to their motion. Note that, within these rules, a LE head will go through a CTCF which is pointing away from it, but would subsequently halt if it diffuses back to the CTCF, as now this would be oriented towards the LE head. LEs (2 heads) attach randomly along the lattice at a rate $k_{\text{on}} = 5 \times 10^{-5}$ step^{-1}, detach at a rate $k_{\text{off}} = k_{\text{on}}$, and like in the 3-D case, if they form a CTCF loop (bring together two convergent CTCFs), they detach at a rate $k_{\text{off}} = 0.1 \times k_{\text{on}}$. The number of LEs is chosen so that the number of attached LEs is roughly 200 throughout the simulation.

5.1.2 Simulation details

The system is initialised from an ideal random walk for the chromatin polymer, and a random distribution of TFs, confined in a cubic box of size 220σ. As mentioned in the previous Section, the simulation box dimensions are chosen so that the system is in the dilute regime. The chromatin polymer is then allowed to relax into a flexible self-avoiding polymer by performing an equilibration run, for $5 \times 10^3 \tau$, where consecutive polymer beads are connected via a harmonic potential and the excluded volume interactions are set by a soft potential:

$$U_{\text{eq}} = \sum_{i=1}^{N-1} U_{\text{harm}}(r_{i,i+1}) + \sum_{i=1}^{N-2} U_{\text{bend}}(r_{i,i+1}, r_{i+1,i+2}) + \\ + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} U_{\text{soft}}(r_{ij}) + \sum_{i=1}^{N} U_{\text{wall, WCA}}(r_{w,i}). \tag{5.1}$$

The persistence length is first set to $l_p = 60\sigma$ in the first $10^3\tau$ in order to boost equilibration: by initially making the chromatin stiff, the polymer conformation quickly changes from a crumpled ideal random walk to a swollen polymer. The persistence length is then set to the realistic value of $l_p = 3\sigma$. The harmonic and soft potential parameters are chosen so as to avoid chain crossings: $K_{\text{harm}} =$

$100k_BT$ and $K_{\text{soft}} = 20k_BT$.

After equilibration, the potentials are changed according to the details described in Chapter 3: bonds between consecutive polymer beads are changed to FENE bonds with $r_0 = 1.6\sigma$ and $K_{\text{FENE}} = 30k_BT/\sigma^2$; excluded volume interactions between polymer beads are now set by the WCA potential; chromatin-protein and protein-protein excluded volume interactions are set by a soft potential with $r_c = 2^{1/6}\sigma$ and $K_{\text{soft}} = 200k_BT$; and chromatin-protein attractive interactions are set by a LJ potential with $r_c = 1.8\sigma$, where ϵ depends on the interaction type (see Sec. 5.1.1.2).

Brownian Dynamics simulations were run using LAMMPS, as described in Chapter 3. Loop extrusion and TF switching algorithms are implemented using external codes coupled to LAMMPS. Specifically, loop extrusion is performed by pausing the MD simulation, then creating/deleting LE bonds according to the rules detailed in Section 5.1.1.3 using an external code, and resuming the simulation, every 250τ. TF switching is performed by randomly selecting a fraction of the TFs (set by the parameter k_{switch}) and changing their state from "on" to "off" (or vice-versa), every 5τ.

In order to map simulation times to real units we match the diffusional properties of the chromatin fibre. For that, the mean squared displacement (MSD) averaged over all simulated chromatin beads is measured (see Sec. 3.2.2), and this is used to fit the time unit to the experimentally measured MSD of fluorescently labelled chromatin loci in yeast cells [48]. The mapping varies slightly between the different models, but for simplicity we fix the time unit to $\tau = 50$ ms in all cases.

Using this mapping, cohesin extrudes 2 beads (60 nm) in $250\tau = 12.5$ s, i.e. with a velocity of $v \simeq 0.3$ μm/min $\simeq$ 30 kbp/min. TF switching occurs at rates ranging between 0 and 0.002 s^{-1}. In the model, the friction due to the solvent was set to $\gamma = 1m/\tau$, which therefore leads to a Brownian timescale $\tau_B = \tau = \tau_{\text{in}} = 50$ ms. Simulations were run for a total of $7 \times 10^5\tau \equiv \sim$10 hours or more. This corresponds to about half of the total duration of interphase in typical human cells [1], allowing us to safely model the folding and organisation of chromosomes after these have decondensed from, and before these have condensed into, the mitotic chromosome configuration (i.e. the chromosome configuration established during cell mitosis). In this study, different MD simulations were performed to investigate the properties of the different models considered: (s)TF,

LE, TF+LE, and sTF+LE model (with/without CTCF proteins). In all cases, 10 repeats of each simulation are considered to account for different polymer conformations and cell-to-cell variability in CTCF binding. The 1-D diffusive extrusion simulations were also run for 10 different repeats, for 20×10^6 steps each.

Table 5.1 summarises the simulation parameters.

$N_{\text{chrom}} = 10,000$ beads	$\sigma_{\text{chrom}} = \sigma_{\text{prot}} = 30$ nm
$M_{\text{Euchr.f}} = 250$ prot	$l_p = 3\sigma = 90$ nm
$M_{\text{HP1}} = 300$ prot	$L_{\text{box}} = 220\sigma$
$M_{\text{PRC}} = 850$ prot	$\gamma_{\text{solvent}} = m/\tau$
TFs	LEFs
Chromatin : protein interaction – LJ	$N_{\text{LEFs}} = 200$
Enh/Prom : Euchr.f – $\epsilon = 5.5k_BT$	Extrusion step – every 250τ
Transc. genes : Euchr.f – $\epsilon = 2k_BT$	$k_{\text{off}} = 0.0167\tau^{-1}$
Heterochromatin : HP1 – $\epsilon = 3k_BT$	$k_{\text{off,convCTCFs}} = 0.00167\tau^{-1}$
Polycomb regions : PRC – $\epsilon = 3k_BT$	
$r_c = 1.8\sigma$	
Switching rate $= [0, 100 \times 10^{-6}]\tau^{-1}$	

Table 5.1 *Simulation parameters for the human chromosome model.*

5.1.3 Data Analysis

5.1.3.1 Calculating Chromatin Interaction Maps

The beginning of the 21st century has been marked by the development of several experimental techniques that allow to probe the conformation of chromosomes. Among these, "Hi-C" [78] has been the most widely employed due to its capability of providing data for genome-wide chromosome interactions.

In summary, Hi-C experiments start by cross-linking the genome so that spatially close DNA loci are joined together. Using restriction enzymes, the unlinked DNA is digested, therefore isolating the pairs of cross-linked DNA loci.

The ends of each pair are then ligated and the cross-links removed. Finally, the ligated pairs are individually aligned with the genome, which allows for the identification of the interacting genomic loci. This last sequencing step can be done simultaneously for all regions of the genome, allowing to build a genome-wide map quantifying how often all different genomic loci interact, within a population of cells [78, 111, 136].

In this thesis, simulated Hi-C interaction maps were generated by recording contacts between chromatin beads whose 3-D separation is $< 5\sigma = 150$ nm. In previous work [14], not much difference was found for 3-D separations in the range $90 - 150$ nm, which is reasonable for cross-linking by formaldehyde. The interaction maps were then averaged over time (every $10^3\tau$ for the last $4 \times 10^5\tau$) and over 10 simulation replicates. The interaction values in the simulated maps range from 0 to 1, therefore when plotting simulated maps side-by-side with the HiC map, the latter was rescaled by a factor of 400 (i.e. the HiC interaction values were divided by 400). HiC data for the GM12878 cell line were obtained at 10 kbp bin resolution, using "square root" normalisation[6], from Ref. [111].

For the 1-D diffusive extrusion simulations,"HiC-like" interaction maps were calculated by considering that two lattice sites (chromatin loci) are in contact if they are occupied by matching LE heads. These maps were then averaged over time (every 400 steps) and over 10 simulation replicates. Note that due to this definition of contacts it is not possible to obtain long-range interaction information.

A quantification of the number of local and non-local interactions was also performed by simply setting a threshold for locality and then calculating the total number of interactions recorded in the map between chromatin beads whose genomic separation is smaller (local interactions) or larger (non-local interactions) than the threshold. The ratio of the number of non-local to local interactions was then plotted as a function of the "locality" threshold. This ratio decreases as the threshold value increases.

Another useful technique to quantify chromosomal interactions is the Chromosome Conformation Capture on Chip (4C) technique. While Hi-C experiments probe the interactions between all genomic loci pairs, 4C experiments are designed to probe the interactions of one pre-selected genomic locus with all other regions

[6]"Square root" normalisation of contact frequencies is a simple procedure to rebalance the HiC matrix so as to have approximately constant sums for rows and columns genome-wide.

of the genome. One can say that the output of 4C experiments is equivalent to extracting one line of the Hi-C map for the pre-selected genomic locus. The advantage of this technique is that it allows for experimentalists to obtain higher resolution interaction profiles, as well as to extract information focused on the region at interest.

Thus, in this thesis, "virtual" 4C profiles were also calculated for the HiC and simulation data. This was done by using the "observed over expected" interaction maps, where the expected value for a given pair of beads is the mean interaction level for all pairs of beads with the same genomic separation. Specifically, each diagonal line in the maps parallel to the main diagonal was first divided by the mean value of interactions in that diagonal. This way all interaction values have the same weight irrespective of their distance to the main diagonal. This facilitates the comparison between experimental and simulated profiles. Then, for the highlighted chromatin locus the number of interactions with all the other loci was calculated.

5.1.3.2 Domain boundary detection in interaction maps

Having calculated the chromatin interaction maps, boundaries between interaction domains were detected using the algorithm depicted in Figure 5.5.

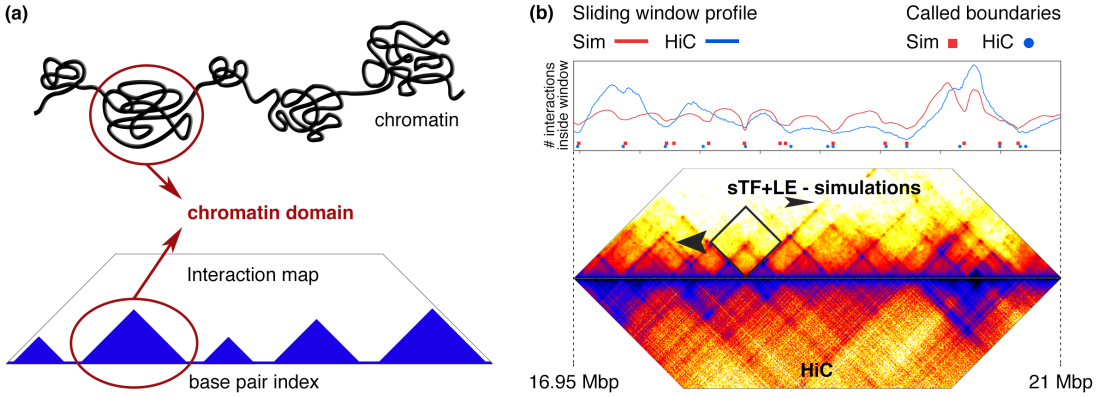

Figure 5.5 *Contact maps are used to record the frequency of interactions between every two chromatin beads. (a) Chromatin domains are defined as regions where there are more intra-domain interactions than inter-domain. (b) Domain boundaries are identified using an algorithm where a window slides along the map diagonal and records the number of interactions inside the window as a function of the genomic position. The algorithm yields an interaction profile with maxima at the centre of domains and minima at the boundaries.*

Specifically, the interaction map is analysed using a square window, of size 300 kbp×300 kbp, that slides along the main diagonal and records the number of interactions inside the window as a function of its genomic position. This algorithm yields a profile of interactions with maxima at the centre of domains and minima at the boundaries, hence a boundary is called for every local minima. The resulting set of boundaries is then verified manually (via visual inspection) to correct wrong boundary calls due to the noisy signal of the interaction maps. The same method was used to call boundaries in Hi-C and simulated contact maps. Boundary locations were said to have been correctly predicted if they appear in the simulated map up to 27 kbp (3 beads) away from the location on the HiC map. To quantify the agreement between called boundaries, the Jaccard index was used, defined as

$$J(B_{\mathrm{HiC}}, B_{\mathrm{sim}}) = \frac{|B_{\mathrm{HiC}} \cap B_{\mathrm{sim}}|}{|B_{\mathrm{HiC}} \cup B_{\mathrm{sim}}|}, \tag{5.2}$$

where B_{HiC} and B_{sim} are the sets of boundaries from the HiC data and simulations respectively. This takes values in the range from zero for no correctly predicted boundaries, to 1 for 100% agreement between simulation and HiC.

5.2 Complementary role of transcription factors and cohesin

In order to characterise the role of transcription factors and loop extruding proteins on the spatial organisation of chromosomes, chromatin folding simulations were first performed by applying the TF model and the LE model separately. Both models were first applied in their simplest forms, i.e., post-translational modification of TFs was not considered so that TFs remain in the "on" state through the whole simulation, and LE proteins were considered to actively extrude chromatin (more details in Secs. 5.1.1.2 (TF model) and 5.1.1.3 (LE model)). Figure 5.6(a-b) shows the chromatin interaction maps for both simulations, comparing the simulated (top triangle) with the Hi-C (bottom triangle) map. The top maps correspond to the full simulated chromosome section, followed by zooms on an inert chromatin region (middle maps) and on a region enriched in active/inactive chromatin (bottom maps).

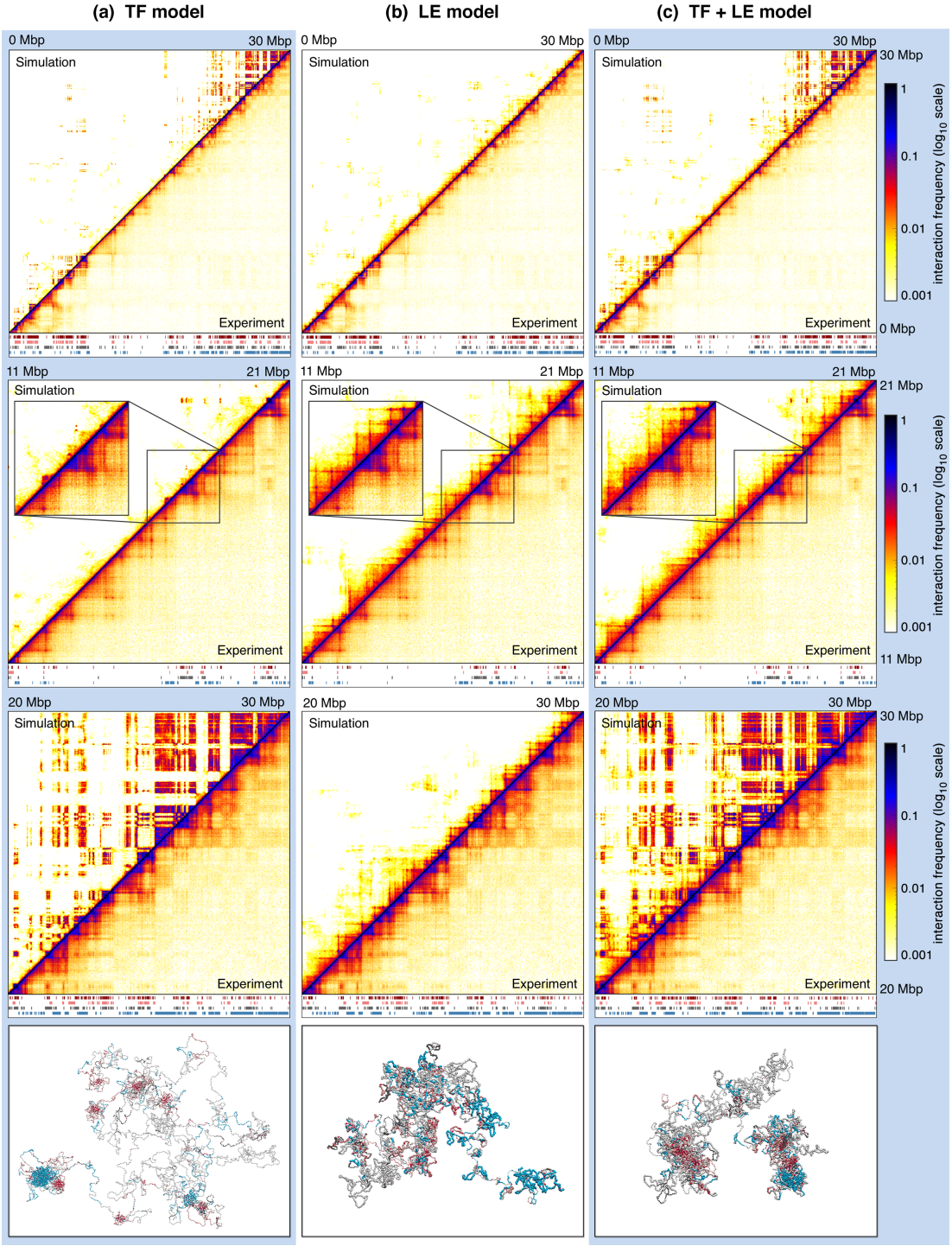

Figure 5.6 *Interaction maps for (a) the TF model, (b) the LE model, and (c) their combination (TF+LE model). Top plots show the full simulated chromosome section (human chr7:1-30,000,000), with zooms on two regions shown below. Underneath each map the considered histone modifications are indicated (see Fig. 5.2 for details). TFs predict the phase separation of active/inactive regions, but fail to predict the folding of chromatin in inert regions. On the contrary, LEs correctly predict the formation of TADs in regions devoid of marks, but are unable to generate interactions between domains bearing the same histone modifications. The bottom panels show snapshots of the simulation for each model.*

In the light of recent simulation studies [14], where TFs binding to distinct chromatin loci were found to self-assemble into distinct segregated clusters, one should expect that the formation of chromatin domains which bear different epigenetic marks will be well captured by the TF model: they phase separate into distinct 3-D compartments, and clusters of like proteins form. Such protein clusters are indeed visible in the simulation snapshot in Figure 5.6(a). Furthermore, by looking at the dynamics of these clusters, they actually resemble clusters formed by heterochromatin [74, 131] and transcription factories self-assembled within euchromatin [27] (see Suppl. Movie **1** [7]).

One way to measure how well the simulated map predicts the HiC data is to simply count the number of correctly predicted domain boundaries (see Sec. 5.1.3.2 for boundary detection details). By analysing the interaction maps in Figure 5.6(a), one sees that the TF model does correctly capture a large fraction of boundaries in active and inactive regions (see the map for the $20-30$ Mbp segment), as well as the pattern of longer-range interactions between segments bearing similar histone marks. These features are a natural consequence of the spatial segregation (or, more precisely, phase separation into domains with self-limiting size [16]) between active and inactive chromatin.

Figure 5.7 quantitatively confirms this domain segregation. For each chromatin bead, the amount of interactions with beads in active, inactive, and mixed domains[8] was determined. The resulting interaction profiles show that chromatin domains tend to interact with other domains of the same type, i.e., bearing the same histone modification marks, confirming the phase segregation of different type domains observed in the experiments (HiC interaction map).

As well as boundaries between active and inactive domains, alternating binding and non-binding chromatin regions can also give rise to boundaries even between two adjacent active (or inactive) domains. However, the TF model clearly fails to capture the folding of the inert chromatin regions (see, e.g., the map for the $11-21$ Mbp segment), which yields a total fraction of correctly predicted domain boundaries of only $\sim 36\%$, through the whole simulated chromosome section.

Compared to the TF model, the LE model (Fig. 5.6(b)) gives a better prediction of local TAD formation, especially within inert chromatin (see

[7] Suppl. Movie 1: https://vimeo.com/288956913

[8] Domains were defined as the regions between the identified domain boundaries, and their type was set according to the most frequent bead type in that region.

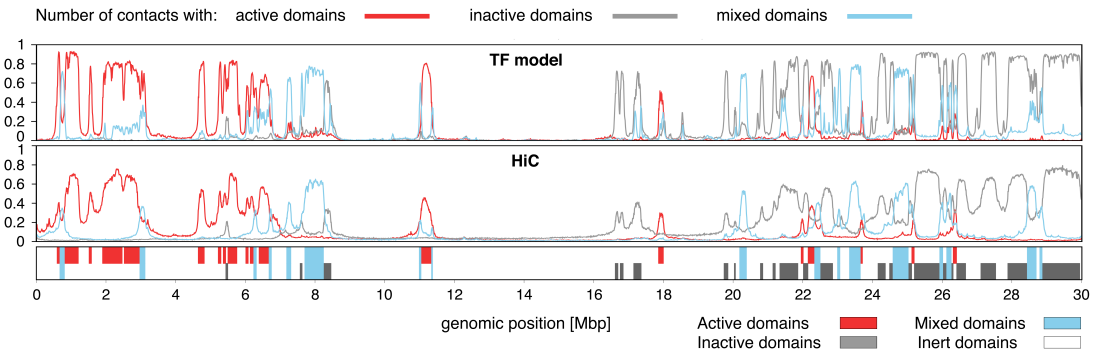

Figure 5.7 *Plot showing, for each chromatin bead, the normalised amount of interactions with active (red), inactive (grey) and mixed (blue) domains. Chromatin domains interact mainly with other domains bearing the same histone marks, a typical feature of the phase separation of active/inactive domains observed in the experiments. For each domain the number of interactions with each type is normalised so that the total number of interactions sums to 1. Domains were defined as the regions between domain boundaries, and their type was set according to the most frequent bead type in that region.*

map for the $11-21$ Mbp segment), where 82% of domain boundaries are correctly predicted. However it performs less well in capturing the higher-order organisation of active and inactive regions (see map for the $20-30$ Mbp segment). Although a high number of boundaries (83%) are correctly predicted in those regions, the interaction maps obtained with the LE model distinctly lack the long-range interactions between domains, which are associated with active/inactive domain phase separation.

The LE model also clearly cannot capture enhancer-promoter interactions within domains unless there are CTCF sites in the vicinity of those regulatory elements. To highlight this, a virtual 4C experiment was "performed" by selecting HiC interactions for a locus corresponding to a promoter, according to the procedure detailed in Section 5.1.3.1 (see Fig. 5.8). The results clearly show that the LE model fails to capture the interaction pattern qualitatively: the correlation between the virtual 4C interaction profiles for the simulated and HiC data is -0.003 for the LE model, compared to 0.27 for the TF model (which is statistically significant, with a corresponding p-value $< 10^{-5}$).

Since each of the models captures different features of chromosome folding, one might expect that a combination of the two should perform much better than either on its own. Indeed the combined TF+LE model yields an improvement, as

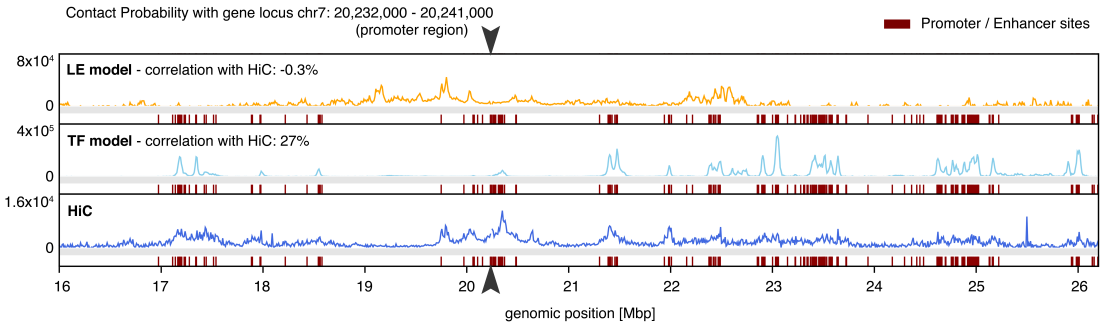

Figure 5.8 *Plots showing interactions with a promoter at position chr7:20,232,000-20,241,000. The black arrowheads indicate the 4C viewpoint. Other enhancer/promoter sites are indicated in red below the curves.*

now both inert and active/inactive regions are in fair qualitative agreement with HiC (Fig. 5.6(c)). An important result is that, within the simulations, extruders do not play a crucial role on local folding within regions that display well-defined patterns of histone modification – their organisation is mainly driven by TF bridging (Fig. 5.6(c), bottom map). For instance, the simulated interaction maps for the TF and TF+LE model in the $20-30$ Mbp region, which is rich in active and inactive domains, are highly correlated (Pearson's correlation $r = 0.76$). The main reason for this is that when bridges bind they tend to compact a whole stretch of chromatin, creating many more contacts compared to extruders, each of which only forms a single loop.

Notwithstanding the improved agreement with HiC, a visual inspection of the interaction maps in Figure 5.6(c) reveals that there are some remaining qualitative discrepancies. Most notably, in the map for the $20-30$ Mbp chromatin region, there are more long-range inter-domain interactions in the simulated interaction maps than in the HiC map, where the number of long-range interactions decays faster with increasing domain-domain genomic distance.

5.3 Switching TFs render better agreement with HiC and FRAP

In order to improve the agreement with HiC experiments [111], a variation of the TF model was taken into consideration. In the TF model discussed in the previous Section, active and inactive factors interact with chromatin beads

thermodynamically. That is to say, there is an attractive binding interaction between the TF and chromatin beads: a TF can bind chromatin when it diffuses into contact, and then unbind due to the thermal motion in the system. The residence time depends on the interaction strength (see Sec. 5.1.1.2) and it is strongly modulated by emergent behaviour such as the "bridging-induced attraction" [17]. More specifically, once a multivalent factor reaches a configuration where it can form multiple chromatin interactions, it remains bound for a long time, as unbinding requires climbing over a potential energy barrier whose height increases with the number of interactions. In the model here presented, typical residence times can encompass the total simulation time, thus the model fails to capture the rapid turn-over of TFs observed *in vivo* (typically of the order of minutes [79, 134]).

Many TFs and other proteins which are relevant to the model are observed in stable foci which also exhibit rapid protein turnover. These two features are difficult to reconcile, but possible explanations are: that protein unbinding can be programmed rather than thermodynamic, as in the case of RNA polymerase II that rapidly unbinds in response to transcription-termination signals [1]; that there is ongoing post-translational modification of proteins which affects binding affinities (e.g. phosphorylation [52, 94]); or that nuclear proteins are actively degraded by the cell and replaced *de novo*. A generic way to model these non-equilibrium unbinding processes is to consider TFs that switch between an "on" (chromatin-binding) and an "off" (non-binding) state at rate k_{switch} (more details in Sec. 5.1.1.2). Recent simulation studies [16] showed that, in simple "toy chromosome" simulations, this switching-TF (sTF) model gives rise to the formation of dynamic protein clusters, reminiscent of nuclear bodies [95], and can affect chromatin interaction patterns.

Therefore, the switching TF (sTF) model was then coupled to the LE model, and its performance in predicting 3D structure in human chromosomes was evaluated. Figure 5.9(a) shows the qualitative effect of TF switching on the interaction maps for different values of the switching rate k_{switch}. The main difference between the combined TF+LE models with and without switching is that switching markedly affects long-range inter-domain, but not intra-domain, interactions: active domains which are far apart along the genome are less likely to interact. This reduces the intensity of the off-diagonal features in the predicted interaction maps, rendering them qualitatively more similar to the HiC. This is

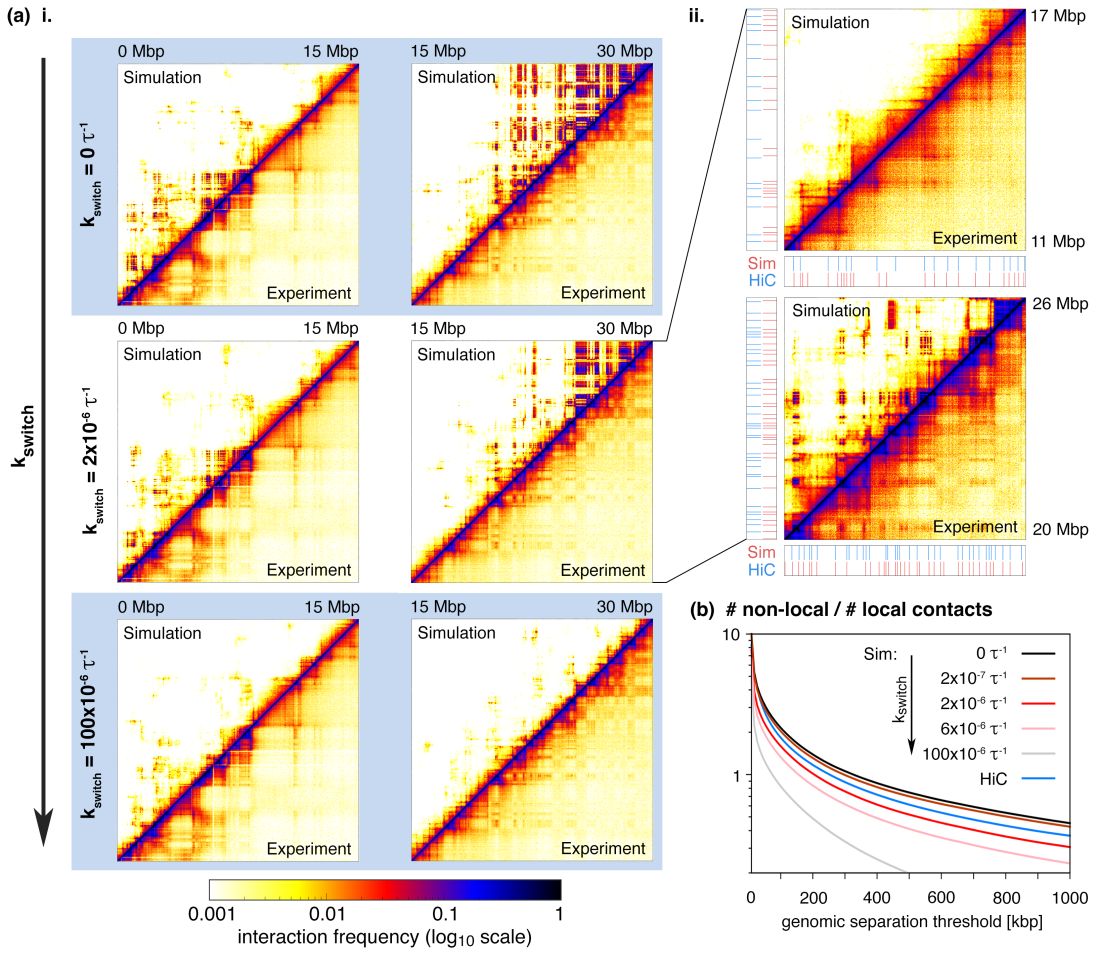

Figure 5.9 *(a) i. Interaction maps comparing sTF+LE simulations with HiC data. From top to bottom the switching rate k_{switch} is increased from 0 to 100×10^{-6} τ^{-1}. Higher switching rates yield less inter-domain interactions. ii. Zooms on different regions of the simulated section with $k_{switch} = 2 \times 10^{-6}$ τ^{-1} (corresponding to $40 \times 10^{-6} s^{-1}$). Below and to the left of the maps the positions of domain boundaries are indicated. The sTF+LE correctly predicts 84% of the HiC domain boundaries. (b) Ratio of non-local to local interactions for HiC and simulation maps, for varying k_{switch} values. Instead of fixing a threshold for locality, we plot the ratio as a function of the threshold for each case. The TF+LE model without switching predicts too high a ratio of non-local to local interactions, when comparing with the HiC curve.*

also shown by the decrease in the ratio of non-local to local interactions with an increasing k_{switch} (Fig. 5.9(b)): switching TFs are needed to predict the correct balance between long- and short-range interactions (see Sec. 5.1.3.1 for details on determining the ratio of non-local to local interactions).

The sTF+LE model also conforms much better with observations from live-cell fluorescence microscopy experiments, which probe dynamical information

inaccessible to HiC. In the absence of switching, the TF dynamics in the model is slow, whereas it is much more rapid with switching (Fig. 5.10, and Suppl. Movies **1** and **2** [9]). Also, the observed macroscopic dynamics of domains formed by non-switching and switching proteins are profoundly different: in the latter case, there are many more events corresponding to clusters splitting and reforming, and also the clusters are smaller. Whilst high-throughput experiments showing the dynamics of chromatin interactions over time are not yet available, the more dynamical picture emergent from the switching model is consistent with fluorescence recovery after photobleaching (FRAP [16]) and single-molecule imaging experiments, which suggest that TF binding is short-lived and lasts for not more than minutes [79, 134]. The differences are clear if one examines the trajectories of individual TFs (Fig. 5.10(a)): without switching, a TF diffuses until it joins a cluster (of like proteins and binding sites), where it tends to stay for the remainder of the simulation; with switching a TF joins a cluster for a short time, then undergoes a period of free diffusion, before joining another cluster. This hopping between clusters leaves a clear signature in the distribution of the mean squared displacement for a given time interval (Fig. 5.10(b)): the distribution for the switching case has a longer tail.

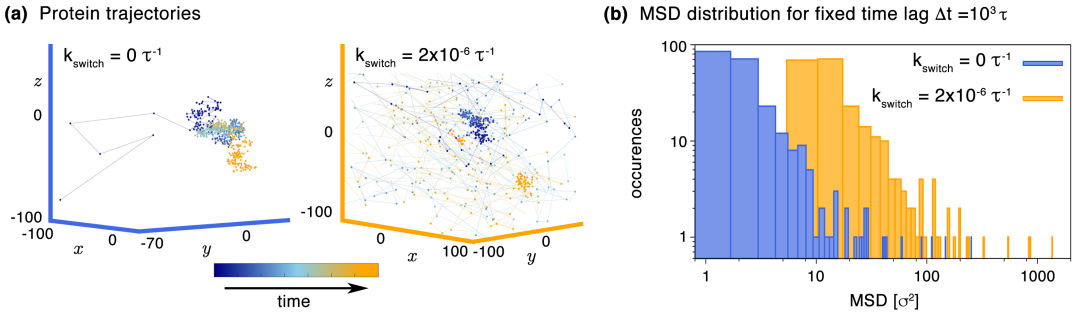

Figure 5.10 *(a) Trajectories in 3-D for a single protein without (left) and with (right) switching. Without switching, the TF quickly gets "stuck" in a stable cluster. With switching, the TF stays in a cluster for a short time, before unbinding and diffusing freely, and later joining a different cluster. (b) Protein dynamics are characterised by the mean squared displacement (MSD) of TFs for a fixed time interval of* $10^3\tau$.

The agreement between the sTF+LE model and HiC can be further quantified, drawing the comparison with the other models considered (see Fig. 5.11). Three quantities are chosen for this analysis: the number of correctly predicted HiC domain boundaries, the correlation between simulated and HiC interaction

[9] Suppl. Movies 1 (https://vimeo.com/288956913) and 2 (https://vimeo.com/288959299)

profiles for long-range chromatin interactions, and the ratio of non-local to local interactions in the simulated and HiC maps.

Starting with the model performance in domain boundary generation (see Fig. 5.11(c)), the sTF+LE model correctly predicts $\sim 84\%$ of the HiC boundaries, across the whole simulated region. This performance is similar to that of the LE and TF+LE models ($\sim$ 83% and $\sim$ 82% respectively), but substantially

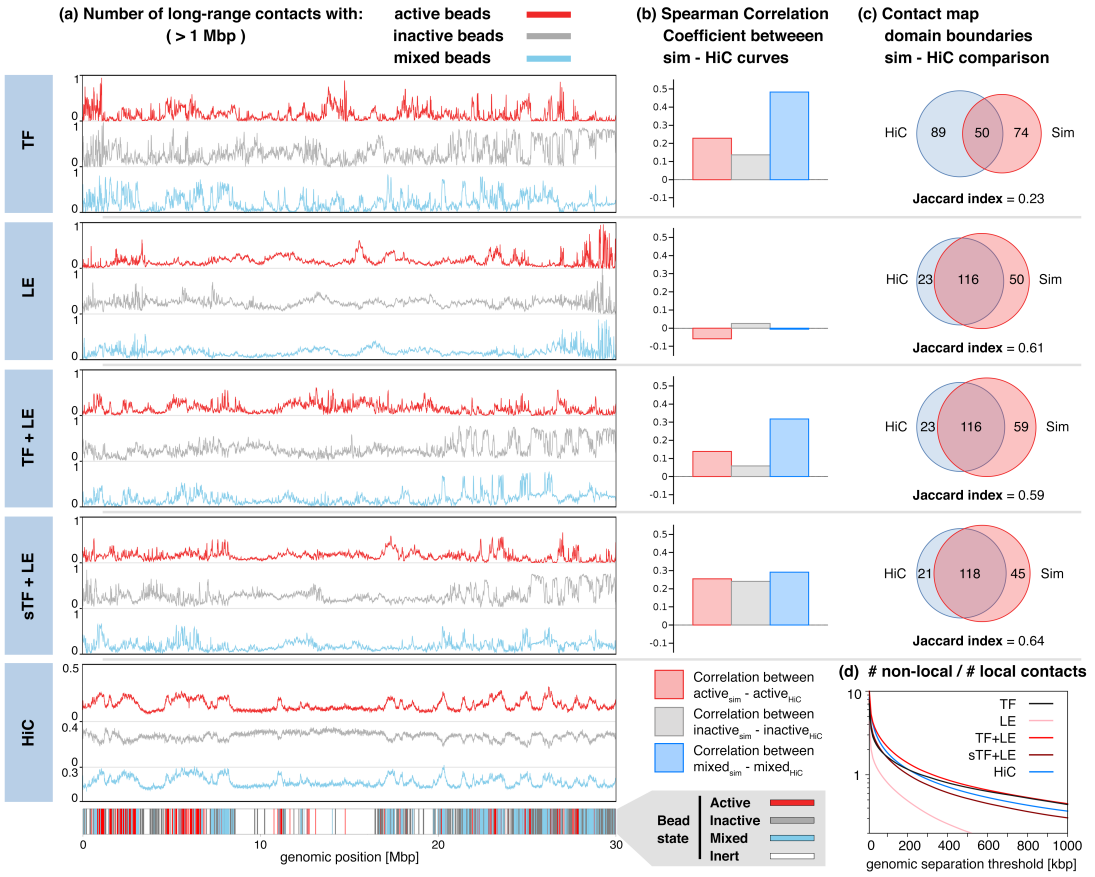

Figure 5.11 *(a) Interaction profiles indicating, for each chromatin bead, the level of long-range interaction with active (red), inactive (grey) or mixed (blue) beads. For each bead the number of interactions with each type is normalised so that the total number of interactions sums to* 1*. (b) Bar plot showing the Spearman rank correlation coefficient for comparison of each curve in (a) with the corresponding HiC data. The LE model yields long-range interaction patterns that essentially do not correlate with HiC. In general, models with TFs give a good correlation with the HiC profiles. (c) Venn diagrams showing the overlap between called domain boundaries in the HiC and simulated contact maps. Since boundaries are related with the fine domain structure near the interaction map's diagonal, models with LEs give in general a good prediction. (d) Ratio of non-local to local interactions for HiC and simulation maps, for the different models, as a function of the locality threshold. The sTF+LE model gives the best prediction of the HiC curve (see also Fig. 5.9(b)).*

better than the TF model ($\sim$ 36%), since the latter fails to predict any of the numerous CTCF-related boundaries within the inert region. Previous studies with only TFs [14], or only extruders [45] found similarly high values, however neither focused on chromosome regions containing both inert and active/inactive regions, as done in this thesis.

Second, to assess how well each model captures domain phase separation and formation of promoter-enhancer hubs [27], one can measure the interaction profiles for long-range interactions between active-active, inactive-inactive, and active-inactive chromosome loci. To do this each chromatin bead is labeled as active or inactive according to whether it binds to active or inactive TFs (which is in turn based on histone modification data). Beads which can bind to both are labeled as "mixed", and those which bind neither as "inert". Figure 5.11(a) shows the fraction of active, inactive and mixed beads which each chromatin bead interacts with, for each of the models. In the HiC profiles there is an enrichment of active-active and inactive-inactive interactions, which is associated with active/inactive domain phase separation. This is captured by models with TFs, as there is a high correlation between the simulated and HiC interaction profiles (see Fig. 5.11(b)). However the LE model fails to capture this long-range interaction pattern, as it shows essentially no correlation with the HiC profiles. This is due to the fact that the LE model fails to predict chromatin interactions beyond the TAD level ($\sim$ 1 Mbp), whose average size in this model depends on the LE processivity and CTCF locations. A Spearman correlation test shows that the combined model with switching performs better than the combined model without switching (although it is not significantly better than the TF only model).

Third, the relative balance between non-local and local interactions was compared for the various models and HiC experiments (Fig. 5.11(d)). This further supports the idea that the TF and LE model separately cannot fully account for HiC data, and also shows that to capture the right decay of the non-local to local fraction a model with switching is required (see also Fig. 5.9(b)).

In summary, the quantitative analysis shows that: chromatin loop extrusion is necessary to correctly reproduce the HiC domain boundaries across the modelled chromatin segment; TF binding is required to reproduce the long-range pattern observed in the HiC interaction map, characteristic of the active/inactive domain phase separation; and TF switching is necessary for the rapid protein turnover in

the protein clusters, which leads to right fraction of non-local to local chromatin interactions in the HiC map. The large-scale interphase chromatin organisation is thus best explained by the sTF+LE model.

5.4 "Diffusive" cohesin also yields domain and CTCF loop formation

Results in Sections 5.2 and 5.3 strongly suggest that chromatin extrusion, regulated by the interaction with CTCF, is fundamental to capture the HiC patterns found especially in inert chromatin regions. The natural question that follows is whether *active* extrusion (where LEs move unidirectionally due to some motor effect as in Refs. [45, 120]) is necessarily required, or whether *diffusive* extruders (dLE) behave similarly. This is currently a relevant question as single molecule experiments on cohesin loaded onto DNA or reconstituted chromatin [33, 65, 130] have not yet found evidence of a direct motor activity. Another plausible alternative is that extruders move via an indirect motor activity, e.g. by a transcribing polymerase, which has been suggested on the basis of simulations [109] and experiments [33]; however it is difficult to find direct evidence for this *in vivo*.

Simulating large chromosome regions with a dLE model in 3-D requires using either infeasibly long simulation times, or using substantially coarser polymer resolutions. Therefore, diffusive extrusion was here studied by using a 1-D lattice model (see Sec. 5.1.1.3 for more details). CTCF sites were positioned as in the 3-D simulations with active LEs, and, like before, it is assumed that the diffusing LEs interact strongly and directionally with CTCFs. "HiC-like" interaction maps can be computed within this 1-D model by considering a pair of monomers (lattice sites) to be in contact if they are "bound" by a diffusing LE, i.e., if the lattice sites are occupied two matching dLE heads (see Sec. 5.1.3.1).

The resulting interaction maps are plotted in Figure 5.12, together with maps from 3-D active LE simulations, computed in the same "1-D fashion" (i. e., two chromatin beads are considered to be in contact if they are bound by a LE). Results show that dLE is essentially indistinguishable from active LE, both visually and quantitatively: 83% of the HiC domain boundaries were correctly predicted by the 1-D dLE model. These results thus show that, even in the

absence of any motor activity, the ability of cohesin to slide diffusively as a molecular slip-link is sufficient to create HiC-like chromatin interaction patterns.

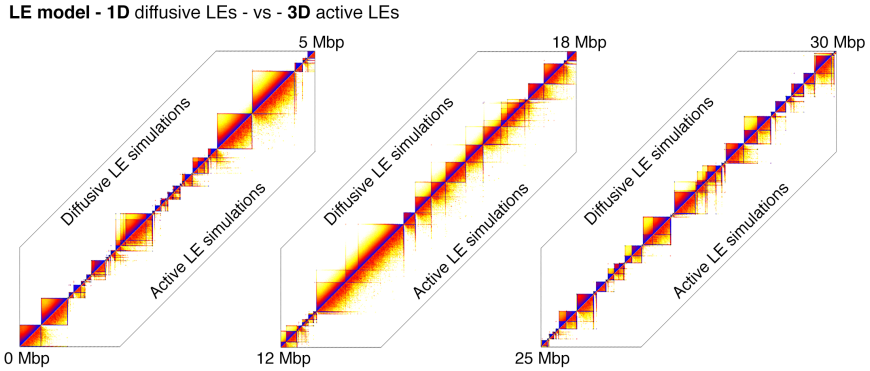

Figure 5.12 *Plots comparing simulated HiC maps generated by the diffusive and active LE model. The active LE results are from the 3-D simulations as shown in previous plots, whereas the dLE results are from a 1-D model. In order to make a fair comparison, here the interaction map for the 3-D simulations was calculated in a "1-D fashion", i. e., two chromatin beads are considered to be in contact if they are bound by a LE. Note that due to the way the interactions are defined it is not possible to generate long-range contacts in the 1-D model.*

5.5 Predicting the effects of protein knock-outs

The combined sTF+LE model proposed in this thesis was further tested by simulating the effect of two different protein knock-outs, which were both recently explored experimentally: cohesin removal and targeted CTCF degradation. Cohesin KO experiments are simulated using the sTF model, i.e., without considering extruders. CTCF KO simulations are performed by omitting CTCF binding sites, so that LEs can freely move until they unbind, or are blocked by other LEs.

The knock-out experiments for cohesin [122] and CTCF [99] were performed in liver and mouse embryonic stem cells respectively, so are not directly comparable to the human cell simulations of this thesis. Nevertheless I will show example interaction maps from these data in Fig. 5.14, for a qualitative comparison. These data sets were also used to generate the plots in Fig. 5.13, showing the ratio of non-local to local interactions as a function of the threshold. As a suitable comparison for the simulations, the first 30 Mb section of mouse chromosome 7 was selected, which has a similar proportion of active/repressed and gene desert regions as the

simulated section. Note that while the plots would differ for different chromosome sections, the trends between the wild type (WT) and knock-out (KO) cases are the same.

Figure 5.13(a) summarises the results for the cohesin KO simulations. Figure 5.13(a)i shows that cohesin removal in the simulations leads to loss of folding in inert chromatin regions (11–21 Mbp), leaving little structure in the interaction maps. This mirrors observations from experiments that knocked out NIPBL, which is required for cohesin loading in mammalian cells [110, 122]. On the other hand, domains organised by active and inactive switching factors (20–30 Mbp) are only subtly affected in the simulations. This is qualitatively consistent with the results of Ref. [122], which found some residual structure in active/inactive regions, but not in inert ones, following cohesin removal in mouse liver cells (see Suppl. Fig. 5 in Ref. [122]). To further illustrate this point, Figure 5.14(a)i shows HiC maps for simulations (left plots) and mouse liver cell experiments for an active region in a similar chromosome region as considered here (right plots). In both cases, the bottom triangle in the map refers to the wild type and the top triangle to the knock-out. From these maps it can be seen that in the mouse HiC data (right plots) the overall contact pattern, as well as some peaks, remain upon NIPBL knock-out, corroborating the simulation results.

Also, like in the experiments, the simulated interaction map reveals stronger active/inactive domain segregation upon cohesin removal: there is a decrease in the number of interactions between domains with different epigenetic marks, and an enhancement of the interactions between like domains (see Fig. 5.15(a)). I should stress that in this work I refrain from a direct comparison of compartmentalisation scores (a measure or active/inactive domain phase separation) in simulations and experiments because in the simulations the nuclear lamina is not directly modelled, which is often associated with inactive regions, and the system conditions are considered to be more dilute than in the nucleus, to avoid non-physical confinement effects.

To better access the qualitative agreement with experiments, the interaction maps were used to extract the ratio of non-local to local interactions as a function of the genomic separation threshold for "locality". Figure. 5.13(a)ii shows the plots comparing the KO and WT cases, for the simulate and HiC maps, obtained from mouse liver cell experiments [122]. There are two distinct regimes: for thresholds below the TAD range ($\sim 700 - 800$ kbp) there is a loss in non-

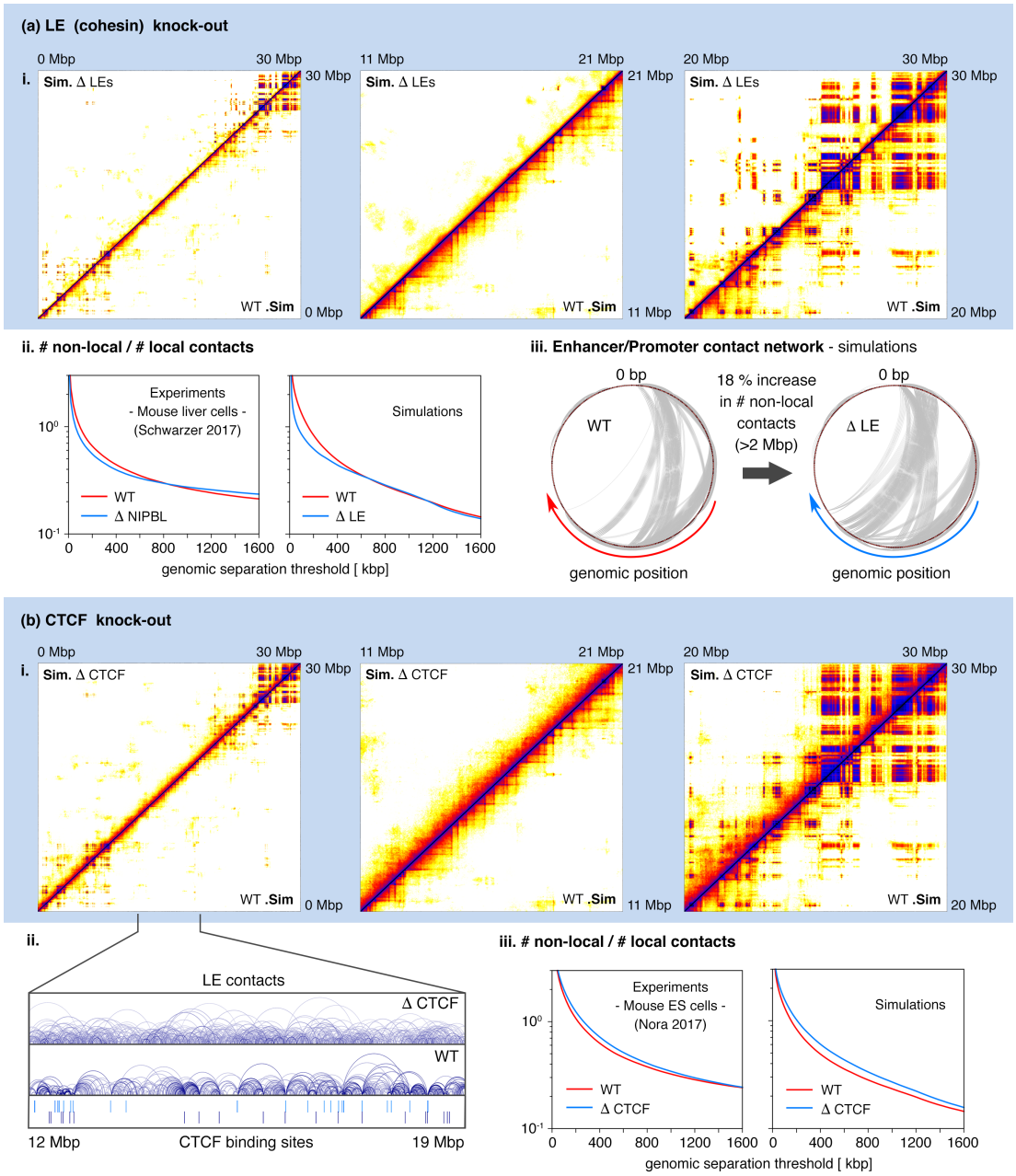

Figure 5.13 *(a) A cohesin knock-out simulation is performed by removing LEs. i. Interaction maps where the knock-out (ΔLEs) is compared with the wild type (WT) simulation. ii. Ratio of non-local to local interactions as a function of the threshold for the KO and WT cases, for simulations (right) and experiments in mouse liver cells (left). iii. Circos plots showing the interaction network between H3K4me1/H3k4me3 chromatin beads, for the simulated KO and WT cases. Interacting beads are connected by grey lines. (b) A CTCF knock-out simulation is performed by omitting CTCF binding sites. i. Similar interaction maps to the above. ii. 1-D representation of LE contacts for the simulated CTCF KO and WT cases. The arc connects chromatin beads bound by a LE. iii. Ratio of non-local to local interactions for the CTCF KO and WT cases, for simulations (right) and experiments in mouse embryonic stem cells (left).*

local interactions upon cohesin removal, and above the TAD range there is a loss in local interactions. The sTF+LE model captures these features up to a threshold $\sim$ 1100 kbp. Above that the simulation predictions deviate from the experimental observations – the WT model yields a higher ratio of non-local to local interactions. This deviation from the experimental observations is due to the choice of low concentrations in the simulations (made to avoid non-physical confinement effects), which allows the polymer to change its conformation faster, meaning that loop extrusion will in fact favour a more compact structure and therefore more non-local interactions (see simulation snapshots in Fig. 5.6(c)).

Cohesin knock-out experiments [110] also reported the formation of superenhancer hubs. Superenhancers are genomic regions containing a high linear density of enhancer elements and high levels of the associated H3K27ac histone modification. Interactions between superenhancers were found to increase after cohesin removal, and examination of HiC ligation events revealed a higher occurrence of triplets of these loci appearing together [9], i.e. three of these loci were in close proximity at the same time. Therefore one can ask whether the knock-out of extruders, in the simulations, leads to an enhancement of interactions between enhancer/promoter chromatin beads. For that, the total number of interactions between such beads was calculated for the WT and KO cases. Figure 5.13(a)iii shows the interaction network for the enhancer/promoter chromatin beads, for the WT and KO cases, in the form of a circos diagram – the chromatin beads are ordered according to their genomic position along the outer circumference in the clock-wise direction and interactions are drawn as grey lines (only interactions with probability > 0.1 are displayed for better visualisation). Upon cohesin KO there is an 18% increase in the number of non-local interactions (genomic separation $>$ 2 Mbp). This result is further supported by analysing clusters of TFs binding active euchromatin, formed through the bridging-induced attraction. These indeed involve more non-local interactions between binding sites after LE removal: the mean genomic separation of chromatin beads associated with such clusters raises by over 10% from 837 kbp to 948 kbp. It is therefore tempting to associate these active protein clusters with the superenhancer hubs found experimentally. The simulations also show that cohesin loss results in a minor decrease in the sizes of TF clusters (however the change is not statistically significant according to a Kolmogorov-Smirnov test).

Fig. 5.13(b) summarises the results for the CTCF KO simulations. These show

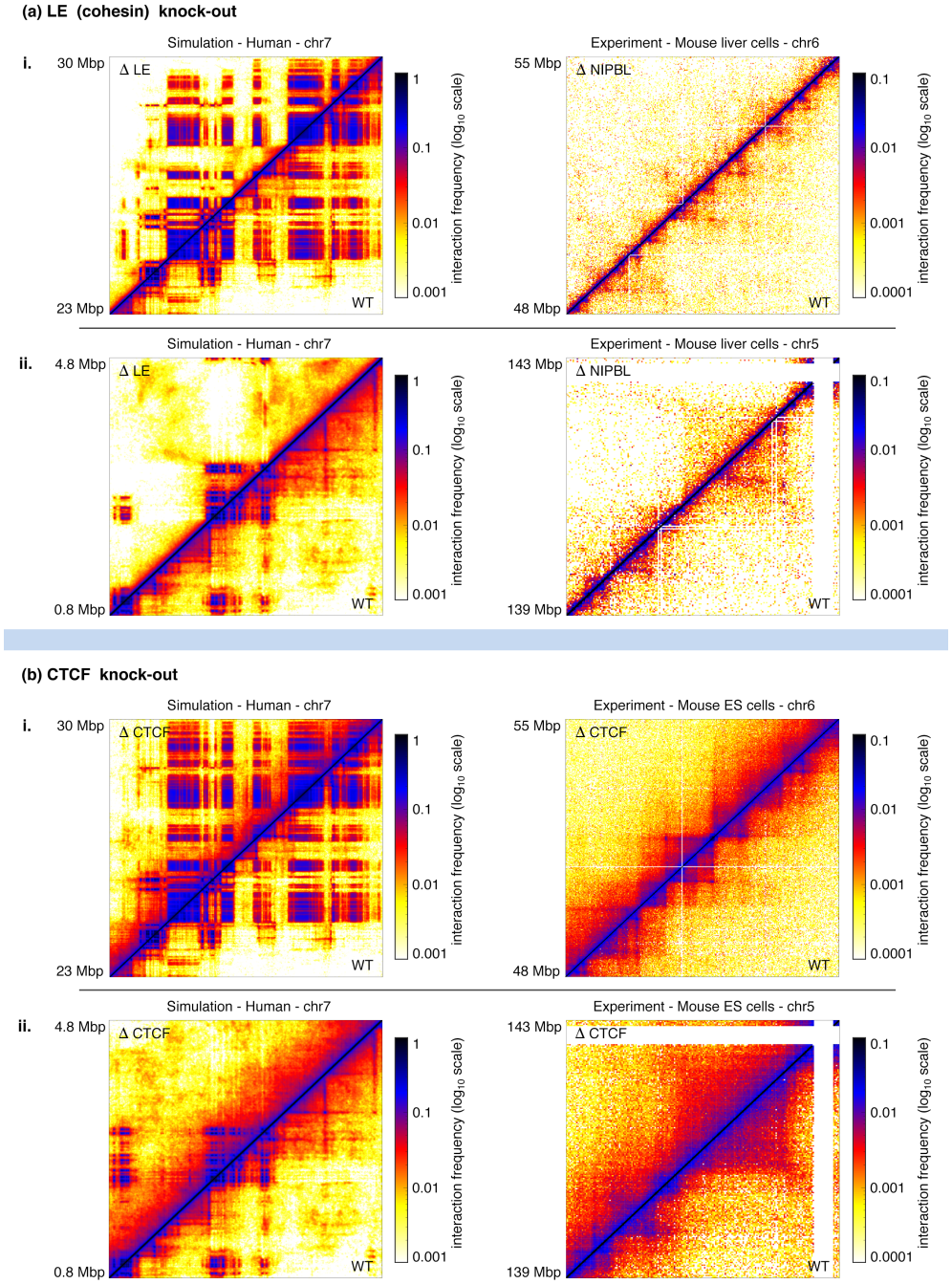

Figure 5.14 *(a) A cohesin knock-out is performed by removing LEs in simulations and the cohesin loader NIPBL in mouse liver cell experiments. Interaction maps are shown for two different chromosome regions - i and ii: (left) simulations' map for a human chromosome 7 region, and (right) HiC map for a syntenic chromosome 6 region of mouse liver cells [122]. For each case, the knock-out is compared with the wild type (WT) map in the same plot: (top left triangle) knock-out, and (bottom right triangle) WT. (b) Similar interaction maps to the above, but for CTCF knock-out simulations (left) and mouse embryonic stem cell experiments (right) [99].*

that CTCF removal leads to a loss of "hot-spots" in the interaction map, which in WT nuclei correspond to convergent CTCF loops (Fig. 5.13(b)i). HiC domains and boundaries become more diffuse/less defined within the inert chromatin region, but are relatively unaffected elsewhere. This is clearly visible from the Venn diagram in Fig. 5.15(b): the number of identified domain boundaries decreases from 162 to 75 upon CTCF KO. Also, the spatial distribution of cohesin on the chromosomes is strongly affected (see Fig. 5.13(b)ii), becoming uniform in the absence of CTCF. These findings are in agreement with experiments knocking out CTCF [71, 99] in mouse embryonic stem cells: the plot in Figure 5.14(b)i-right shows that HiC domains are only subtly affected upon CTCF KO in active regions, whereas Figure 5.14(b)ii-right shows that in more inert regions the HiC boundaries become diffuse. However, there is one difference between simulations and experiments. In the WT simulations, cohesin localises mostly at CTCF sites, consistent with ChIP-seq data [20], whereas in the CTCF KO simulations cohesin is distributed uniformly across the chromosome segment. In experiments, cohesin instead accumulates at transcription start-sites upon CTCF loss. One possible reason for this discrepancy is that in the simulations extruders bind to the chromatin fibre at random sites, whereas in real cell systems cohesin might have preferred loading sites on the chromatin: some preferential binding of NIPBL, required for loading, has been observed at transcription start sites [20]. Recent simulation studies [15] show that including preferred loading sites in simulations would in fact, in the absence of CTCF, lead to an enrichment of cohesin at those sites.

Fig. 5.13(b)iii also shows the ratio of non-local to local interactions as a function of the genomic separation threshold for the KO and WT cases, for the simulated and HiC maps, obtained from mouse embryonic stem cell experiments [99]. For all analysed contact separation threshold values, simulations qualitatively agree with the experimental observations: CTCF removal slightly favours more non-local interactions. Such is due to the fact that upon CTCF KO, LEs/cohesin can extrude farther.

Finally, it is worth highlighting here a recent simulation work [100] which is broadly related to the work presented in this Chapter. In that study the loop extrusion model was combined with a block copolymer model [55], which postulates a weak direct attractive interaction between all inactive regions. Whilst this related work also found that both components of the model are

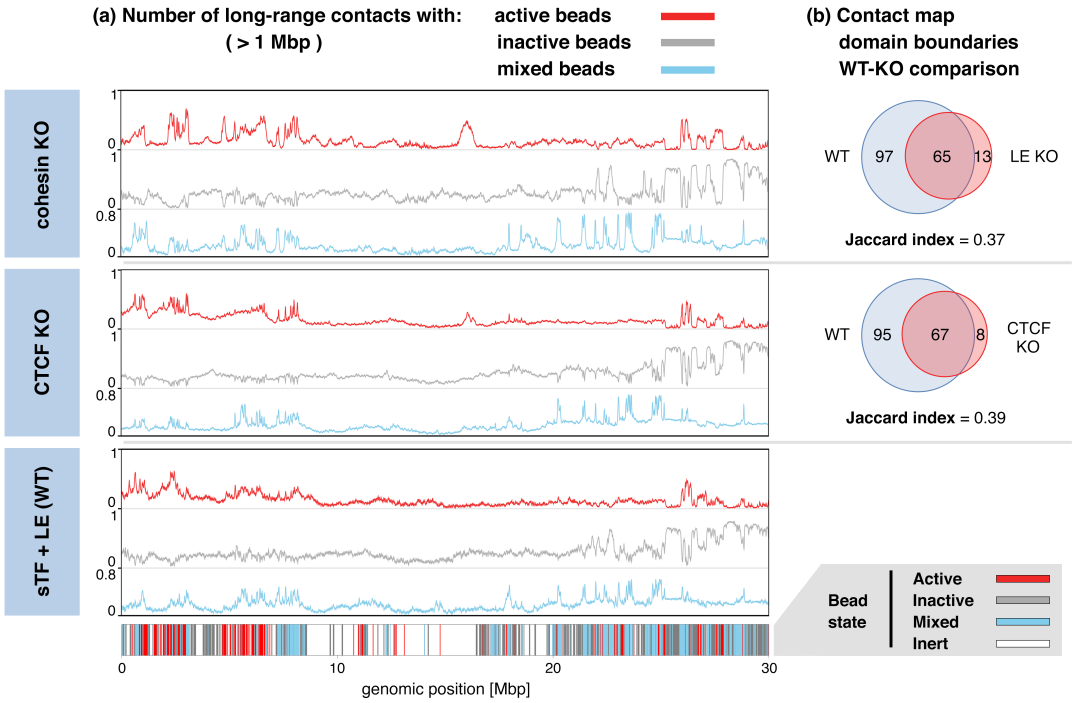

Figure 5.15 *As in Figure. 5.11: (a) Interaction profiles indicating the normalised amount of long-range interactions with active (red), inactive (grey) and mixed (blue) beads, for each chromatin bead. (b) Venn diagrams showing the overlap between called domain boundaries in the KO and WT models.*

required to get good agreement with HiC data, it was suggested there that extrusion may compete against active/inactive domain segregation – e.g., if a convergent CTCF loop spans domains belonging to regions of different types. This interference mechanism is appealing because it is consistent with the observation that cohesin or CTCF removal leads to an enhancement of non-local active/inactive domain phase separation [122]. The work here described, though, does not present evidence of significant competition between chromatin-state and cohesin-mediated folding at a local level. For example, there is little difference between the TF and TF+LE models in the 20–30 Mbp region – the LEs do not interfere with the ability of TFs to organize active/repressed regions. Similarly there is no significant change in the LE loop length distribution between the LE and TF+LE models in either inert or active/repressed regions – the TFs do not interfere with extrusion.

5.6 Remarks

In conclusion, in this Chapter I have conciliated two previously competing viewpoints regarding the mechanisms driving mammalian genome organisation. By means of Brownian dynamics simulations, I singled-out the role of two popular and successful models on the folding of human chromosomes: the transcription factor [7, 14, 17, 98] and loop extrusion [45] models. The TF model is motivated by the abundance of multivalent architectural chromatin-binding proteins or complexes (e.g., HP1, PRC1, TF/PolII complexes etc.), which are known to form loops within the genome, and organise it into active and inactive regions. The TF model naturally explains the observations of transcription factories [27] and nuclear bodies [16] as multivalent TFs generically cluster through the bridging-induced attraction [17]. The LE model is motivated by the evidence that cohesin mediates chromatin looping between convergent CTCF sites in the genomes of mammals [111].

In Section 5.2, I presented simulations of chromosome folding by TFs and LEs, and showed that the results strongly suggest that TFs and cohesin (modelled as LEs) play complementary roles in genome organisation. On one hand, cohesin is necessary to organise and compact regions of inert chromatin (regions void of genes) where depletion of most histone marks is consistent with minimal TF binding. Accordingly, cohesin is required to account for many of the TAD boundaries in the simulated region of human chromosome 7, which contains a large inert chromatin region. On the other hand, activating and repressive TFs are sufficient to organise active and repressed regions respectively, as knocking out extrusion leaves largely similar contact patterns in these regions (Fig. 5.13).

Importantly, in Section 5.4 I showed that an active mechanism for extrusion is not the only model which can generate TADs within inert chromatin: a similar number of HiC boundaries are correctly predicted by a diffusive LE model where cohesin slides along chromatin with no preferred direction (Fig. 5.12). This is a very interesting result since to date there is no direct evidence of unidirectional motion of cohesin on chromatin [33, 65, 130].

In Section 5.3 I proposed a variation of the model that gives the best concordance between simulations and the available experimental data, which includes a biochemical "switching" reaction for TFs, between an "on", chromatin-binding, and an "off", non-binding, state. This reaction drives the system

away from thermodynamic equilibrium, and allows TFs to bind strongly, and yet be able to dissociate frequently. This switching model gives a better prediction of long-range contacts, which would otherwise decay too slowly. More importantly, switching is necessary to reconcile simulations with fluorescence microscopy experiments which measure fast dynamics for both transcription factors [26, 70, 79, 134] and other protein clusters [16].

The combined sTF+LE model was shown to reproduce qualitatively the effect of recent knock-out experiments, in Section 5.5. Cohesin degradation leads to unfolding and the disappearance of domain boundaries in inert chromatin regions, but results in smaller changes within active/inactive chromatin [122]. CTCF knock-out also mainly affects inert chromatin regions, and homogenises the distribution of cohesin along the chromatin fibre [71, 99]. Even though, in the simulations, the effect of TFs and LEs on each other is subtle regarding HiC domain formation, this effect becomes more pronounced with respect to longer-ranged interactions. For instance, experiments showed that cohesin loss leads to the formation of hubs of superenhancers involving very long-range contacts [110], which is associated with the increase in active/inactive domain segregation. This result agress with the model here presented, as protein-mediated interactions between active chromatin beads associated with promoter or enhancers become longer-ranged upon LE knockout.

I shall note, though, that a limitation of the current work is that its methodology relies on previous knowledge of the TFs responsible for folding. A recent approach [11] has introduced a possible way to circumvent this problem, by using polymer physics and machine learning to infer the optimal, minimal number and type of TFs required to reproduce the HiC interaction matrix within a given accuracy. However, this approach requires the HiC data as an input, whereas the work in this thesis relies on a predictive bottom-up approach.

In summary, the results described in this Chapter suggest that transcription factors and cohesin complexes provide two complementary mechanisms for chromosome organisation, and that they are important in different regions of the genome. The question of how this "division of labour" is functionally relevant remains open: one could speculate that cohesin-mediated folding of inert chromatin may be useful to facilitate the transition to mitosis, where (condensin-associated) loops are likely much more abundant. Also, one should stress that the current work focuses on a single cell type during interphase, where histone

modification patterns are already established. It remains possible that LEs and TFs may have a more complex relationship in situations where the underlying epigenetic landscape is dynamic, like during differentiation, or on exit from mitosis [92, 93].

6 Large-scale nuclear organisation of fruit fly chromosomes

In the previous Chapter, the role of different types of proteins on the spatial organisation of chromatin during interphase in human cells was investigated, by comparing the late time simulation results with data from chromosome conformation capture (HiC) techniques.

At the same time as protein-chromatin interactions shape chromatin architecture, many of the regulating proteins are found to also organise into subnuclear structures, forming protein clusters and foci [25, 126]. For instance, transcription factors aggregate into "factories" that regulate transcription of active genes [102], and Polycomb group (PcG) proteins organise into hundreds of nano-clusters (known as Polycomb bodies) that regulate gene repression by stabilising long-range chromatin interactions [138]. Understanding the intimate relation between protein organisation and genome architecture is, thus, a fundamental question in biology, since both have implications on gene regulation and cellular integrity [23, 28, 126].

Despite the number of studies probing chromatin organisation [78, 96, 124] and the properties of protein nuclear clusters [25, 83, 138], it is not yet well understood how both chromatin and proteins reorganise in the nucleus after cell division. Specifically, when cells enter into mitosis, chromosomes are reconfigured into very compact structures. At this point, transcription is shut down and many of the proteins that regulate interphase organisation dissociate from chromatin. As a result, interphase features such as chromatin organisation into topologically associated domains (TADs) and phase segregation of active and inactive chromatin domains (see Ch. 5 for details) are lost during cell division [96].

How then interphase organisation is reestablished remains an open question. One possible mechanism relies on persistent proteins, that remain bound to chromatin during cell division, which then act as post-mitosis nucleation points for the reestablishment of protein clusters [43]. However, the kinetic pathway leading cells from the mitotic to the interphase state remains elusive.

Therefore, the aim of the work described in the present chapter is to give new insights into this question by performing large-scale molecular dynamics simulations of a whole cell nucleus. For simplicity, I will model the nucleus of an haploid *Drosophila melanogaster* cell, i.e. a cell containing only one set of chromosomes, so that the system considered will consist of four fruit fly chromosomes (see Fig. 6.1(a)) and associated proteins, confined in a spherical nucleus.

In Section 6.2, I will start by investigating how the large-scale chromosome conformation evolves, after cell division. I will show that chromosomes relax from the mitotic state (i.e. the condensed chromosome configuration established during cell mitosis) to less compact ellipsoidal structures, preserving the nuclear organisation into discrete territories. The resulting chromatin interaction maps are in good agreement with HiC data, suggesting that large-scale interphase organisation might emerge simply as a result of the post-mitotic configuration and the semi-dilute nuclear conditions.

In Section 6.3, I will show that, at the same time as chromatin reorganises, protein clusters emerge, much resembling nuclear bodies. Indeed, the model correctly predicts the cluster size distribution of PcG bodies observed experimentally [138]. The fact that these clusters form due to protein-chromatin interactions and that the formation of such clusters promotes further chromatin-chromatin interactions suggests that chromosome and protein reorganisation, after cell division, are two intimately linked processes.

In Section 6.4, I will look deeper into the nuclear distribution of chromatin and nuclear bodies. I will show that heterochromatin factors accumulate at the centromeric regions, i.e. in the chromatin regions that constitute the chromosome centromere, where the density of heterochromatin is high [41]. On the other hand, PcG bodies tend to be positioned near open chromatin regions and away from the nuclear centre and periphery, in accordance with experimental observations [25]. The results also show that active and inactive chromatin regions have a similar radial nuclear distribution, whereas silenced chromatin regions are distributed

towards the nuclear periphery due to their attraction with the nuclear lamina. Interestingly the results suggest that, in haploid cells, the concentration of particles in the centre of the nucleus is slightly lower. One explanation for this might be that the interaction between silenced chromatin regions and the nuclear lamina, even though being weak, significantly affects the overall organisation of the chromosomes, "pulling" them towards the nuclear periphery.

Finally in Section 6.5, I will explore in more detail the effect of the interaction between silenced chromatin regions and the nuclear lamina. I will show that in the absence of such interaction the effective particle concentration towards the interior of the nucleus increases, resulting in enhanced chromatin interactions, faster protein cluster growth, and slightly larger protein clusters. This is also reflected in the chromatin nuclear distribution, where active chromatin becomes distributed more towards the nuclear centre and inert (silenced) chromatin becomes evenly distributed across the nucleus, upon turning off chromatin-lamina interactions.

6.1 Modelling fruit fly chromosomes in a spherical nucleus

6.1.1 The system

In this work, the nuclear organisation in *Drosophila melanogaster* haploid cells is studied by modelling the entire chromosomes 2, 3, 4 and X (see Fig. 6.1(a)) confined within a spherical wall in the presence of explicit bridging proteins. Chromatin loop extruding complexes are not here modelled since these have been shown not to have such an important effect on chromatin organisation in *Drosophila* as in mammals [117]. Specifically, the anchor regions of the observed loop domains in *Drosophila* HiC maps were found not to be enriched in CTCF proteins, which together with cohesin are thought to be key organisers of mammalian chromatin loops [111].

6.1.1.1 Chromosomes

Each chromosome arm is modelled, at the resolution of the chromatin fibre, as a self-avoiding semi-flexible polymer composed of spherical monomers, according

to the details described in Chapter 3. Each chromatin bead, with diameter $\sigma = 30$ nm, represents about 3 kbp of DNA, being the chromatin fibre's persistence length set to $l_p = 3\sigma = 90$ nm.

Polymer beads are "coloured" according to their underlying chromatin state based on histone modifications, previously inferred via a hidden Markov model (HMM) that considers multiple ChIP-seq data and explicitly models the presence or absence of each modification along the genome [66] (see Fig. 6.1(b)). Specifically, polymer beads are annotated as corresponding to euchromatin (transcriptionally active enhancers and promoters, and transcribed genes), heterochromatin (repressed/inactive genes), and inert chromatin (silenced genes, i.e. regions of low or no transcriptional activity, also known as black chromatin).

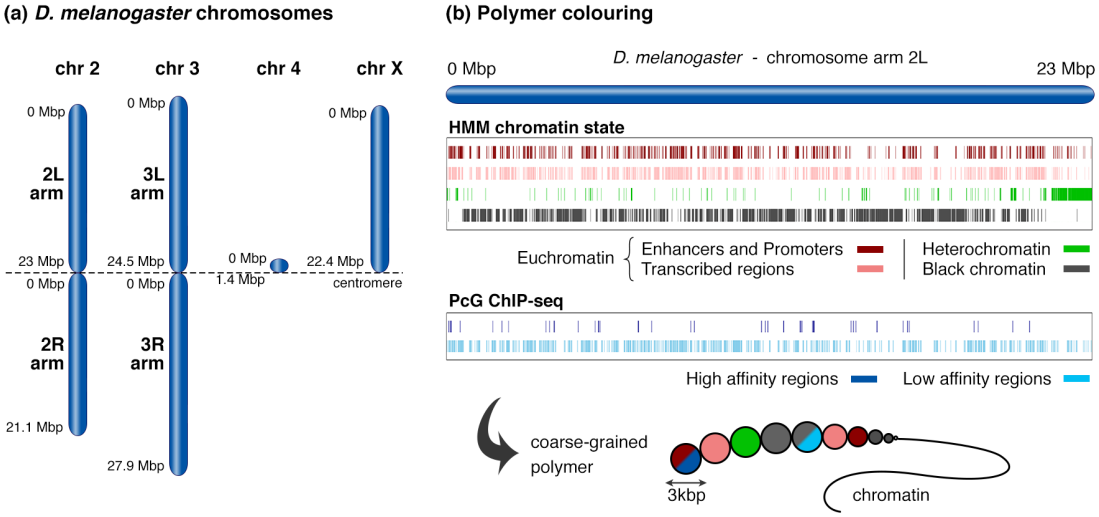

Figure 6.1 *D. melanogaster genome model. (a) The nucleus of an haploid cell is modelled containing one set of chromosomes 2, 3, 4 and X. (b) Chromatin polymer beads (3 kbp) are "coloured" according to the abundance, in the respective chromatin locus, of the Hidden Markov Model (HMM) states, obtained from Ref. [66]. Chromatin beads are thus annotated as euchromatin (enhancers/promoters and transcribed regions), heterochromatin, and inert chromatin (regions of low transcriptional activity and void of most histone modifications, also known as black chromatin). The binding sites for high and low affinity polycomb group (PcG) proteins are also considered – ChIP-seq data are obtained from Ref. [138]. Therefore beads can have two annotations as illustrated in the bottom polymer diagram.*

In addition to these states, the binding sites for high and low affinity polycomb group (PcG) proteins are also considered. For this, ChIP-seq data are directly used to identify the genomic PcG binding locations [138]. Therefore, polymer beads can have two annotations as illustrated in the bottom diagram

of Figure 6.1(b). For instance, a polymer bead can simultaneously represent a transcribed gene and a PcG binding region. But again, as clarified in Chapter 5, this does not mean that the same DNA sequence codes for a gene and a PcG binding site, but that the polymer bead in the model corresponds to a chromatin region bearing DNA sequences that code for one and another. Finally, regions of the genome that are not enriched in any particular mark are defined as "neutral", and left unmarked.

Note that the current model considers a fixed interphase epigenetic landscape, for all chromosomes during the whole simulation. It does not consider the epigenetic changes that occur right after mitosis while the interphase landscape is being reestablished. This choice was made for simplicity, given that this Chapter focuses on the overall late-time nuclear reorganisation of chromosomes and nuclear bodies when the cell goes into interphase.

6.1.1.2 Proteins

Proteins are modelled, for simplicity, as multivalent spherical beads with diameter $\sigma_{\text{prot}} = \sigma$. These interact sterically with each other via the WCA potential in Eq. (3.5). Four types of proteins (or factors) are considered: (i) euchromatin-binding, such as polymerases or CTCF; (ii) heterochromatin-binding, such as HP1; (iii) polycomb-group proteins, such as PRC1; and (iv) non-chromatin binding, which simply act as molecular crowders. A total number of $M = 11000$ proteins is considered, which are split into the four types as follows: (i) 1000, (ii) 2500, (iii) 2500, and (iv) 5000. The number of proteins of each type was chosen according to the relative proportion of euchromatin, heterochromatin, polycomb-binding, and black chromatin polymer beads.

The different proteins bind to their cognate chromatin beads through an attractive interaction set by the LJ potential in Eq. (3.7), with $r_{\text{thr}} = 1.8\sigma$. Euchromatin factors bind strongly, $\epsilon = 6k_BT$, to enhancers and promoters, and weakly, $\epsilon = 3k_BT$, to transcribed regions. Heterochromatin factors bind weakly, $\epsilon = 3k_BT$, to heterochromatin beads. PcG proteins bind strongly, $\epsilon = 6k_BT$, to high-affinity, and weakly, $\epsilon = 3k_BT$, to low-affinity chromatin beads.

Since proteins are modelled as multivalent beads, during the course of a simulation, chromatin-binding factors can bind to multiple chromatin loci bearing the same state annotations, effectively forming bridges between distant loci along

the polymer. Thus, proteins of the same type will naturally cluster through the previously reviewed "'bridging-induced attraction" mechanism (see Sec. 2.3.2), forming nuclear bodies.

Note that the current model for transcription factors that bind to euchromatin, such as polymerases, does not consider explicitly the transcription process where factors/proteins unwind the DNA and process along a single strand in order to transcribe a DNA sequence. Instead, proteins simply diffuse through the nucleus and bind/unbind thermodynamically to the chromosomes once they find their cognate chromatin sites, as per the argument presented in page 59.

6.1.1.3 Nuclear wall

The nuclear wall is modelled as a spherical surface that confines all chromosomes and proteins. All particles interact with the wall through the soft indentation potential in Eq. (3.9), with $K_{\text{indent}} = 200k_BT$.

The size of the nucleus is set so as to have the desired particle volume fraction. For this work a volume fraction of 8% is chosen, meaning that for a total number of modelled particles of 51120 (40120 polymer beads and 11000 proteins) the radius of the spherical nucleus is $\sim 43\sigma \equiv \sim 1.3\mu$m. The chosen particle volume fraction is about half of the realistic value ∼15%. This is reasonable here as we are only modelling one set of chromosomes (haploid cell).

The inner face of the nuclear membrane of most cells is coated with a complex network of intermediate filaments and membrane associated proteins. Together these form the so-called nuclear lamina [1]. Besides providing mechanical support to the nuclear wall, the lamina also plays a key role on chromatin organisation and gene expression regulation: genes located near the lamina remain repressed, being expressed at very low levels [107], and genes can actually be repressed by being repositioned closer to the lamina [113]. In *Drosophila*, almost half of the genome is largely void of most histone marks, being rarely transcribed. These silenced, or inert, regions are known as "black chromatin". Given that peripheral chromatin is generally silenced [28], it has been suggested that transcriptional repression in black chromatin is related to direct interactions with the nuclear lamina [41].

Therefore, the model here presented considers a further interaction between black chromatin and the nuclear lamina. For the purpose of simplicity, the lamina

is not explicitly modelled. Instead, black chromatin interacts attractively with the nuclear wall through the 1-body form of the LJ potential in Eq. (3.7), with $\epsilon = 2k_BT$ and $r_{\text{thr}} = 1.8\sigma$.

6.1.2 Simulation details

In *Drosophila* cells, chromosomes organise in the nucleus according to the so-called Rabl-orientation: the chromosome centromeres cluster near one pole of the nucleus while the ends/telomeres extend towards the opposite pole, so that chromosome arms are more or less parallel to each other [29]. This orientation results from the mechanism of chromosome segregation during mitosis (anaphase), and persists through interphase [50, 80].

Therefore, the system is initialised from a mitotic-like configuration for the chromosomes (see Fig. 6.2), and a random distribution of proteins, confined in a cubic box of size $\sim 140\sigma$. The chromosome centromere beads are tethered to an additional bead fixed to the nuclear wall (large black bead in Fig. 6.2), so that chromosomes display a Rabl orientation. Thus, the last bead of chromosomes 2L, 3L, 4, X and the first bead of chromosomes 2R, 3R interact attractively with the same tethering bead through a soft harmonic potential, with $K_{\text{harm}} = 10k_BT$.

The mitotic conformation is modelled following the approach described in Ref. [115]. The chromosome polymer is set to describe a generalised helix, consisting of a stack of rosettes with 12 "leaves" or loops (see Fig. 6.2(a)). Each rosette leaf is composed by 25 polymer beads, meaning that each rosette has a contour length of $12 \times 25\sigma \approx 900$ kbp. The equations describing the polymer helix are

$$\begin{aligned} x(\phi) &= r_{\text{chr}} \left[f + (1-f)\cos(k\phi)^2 \cos\phi \right] \\ y(\phi) &= r_{\text{chr}} \left[f + (1-f)\cos(k\phi)^2 \sin\phi \right] \\ z(\phi) &= \frac{p\phi}{2\pi} \end{aligned} \tag{6.1}$$

where $f \times r_{\text{chr}} = 0.3 \times 21\sigma \approx 190$ nm is the radius of the mitotic cylinder, $p = \sigma$ is the vertical step for each full turn (i.e. the vertical spacing between rosettes), and $k = 6$ so that each rosette has 12 leaves.

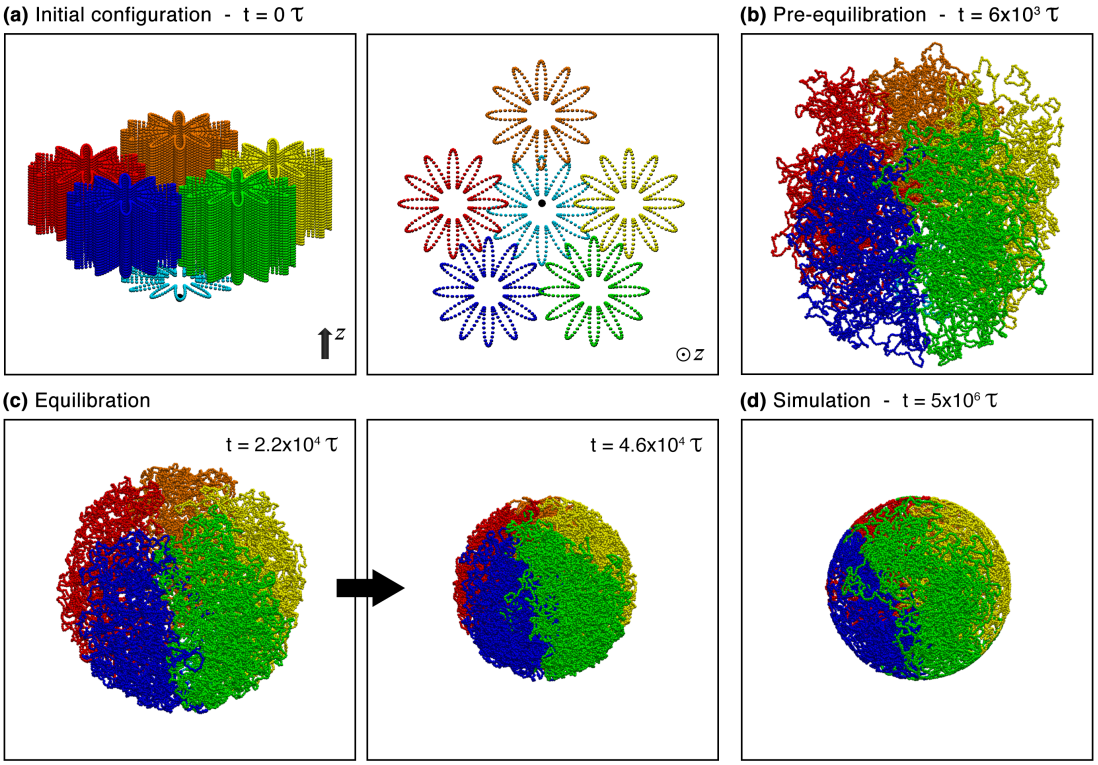

Figure 6.2 *Snapshots showing all chromosomes at different stages of the simulation. (a) At the start of the simulation, chromosomes are initialised in a mitotic cylinder conformation. (b) During pre-equilibration chromosomes relax into self-avoiding semi-flexible polymers. (c) During equilibration chromosomes are confined in a spherical nucleus. (d) After equilibration, chromosomes further relax and interact.*

A pre-equilibration run of $\sim 1 \times 10^4 \tau$ is performed so that the chromosomes are allowed to partially relax into self-avoiding semi-flexible polymers (see Fig. 6.2(b)). During this short run consecutive chromosome beads are connected via a harmonic potential and the excluded volume interactions are set by a soft potential:

$$U_{\text{eq, chr}} = \sum_{i=1}^{N-1} U_{\text{harm}}(r_{i,i+1}) + \sum_{i=1}^{N-2} U_{\text{bend}}(r_{i,i+1}, r_{i+1,i+2}) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} U_{\text{soft}}(r_{ij}) + \sum_{i=1}^{N} U_{\text{wall, WCA}}(r_{w,i}). \tag{6.2}$$

The persistence length is gradually increased from 1σ to 3σ, and the harmonic and soft potential parameters are chosen so as to initially allow for chain

crossings: $K_{\text{harm}} = K_{\text{soft}} = 10k_BT$. This helps to disentangle knotted chromatin conformations that can easily emerge in the beginning of the simulation, close to the mitotic state. During pre-equilibration, all proteins interact sterically with the chromosomes.

Afterwards, another simulation run, this time of $\sim 4 \times 10^4\tau$, is performed to gradually confine the whole system, chromosomes and proteins, in a spherical nucleus, with the desired radius of $\sim 43\sigma$ (see Fig. 6.2(c)).

After the nuclear spherical wall has reached the right size and the chromosomes have partially relaxed, the simulation potentials are changed according to the details described in Chapter 3: bonds between consecutive chromosome beads are changed to FENE bonds with $r_0 = 1.6\sigma$ and $K_{\text{FENE}} = 30k_BT/\sigma^2$; excluded volume interactions between contiguous chromosome beads are set by the WCA potential; all other excluded volume interactions are set by a soft potential with $r_c = 2^{1/6}\sigma$ and $K_{\text{soft}} = 50k_BT$; and interactions between chromosomes and chromatin-binding proteins are now "turned on", being set by a LJ potential with $r_c = 1.8\sigma$ and with ϵ depending on the interaction type (see Sec. 6.1.1.2). With this choice of potential parameters, polymer chain crossing are no longer allowed.

Simulations are then run for a total of 2×10^6 simulation time units (τ) or more. In this work, different versions of the model will be considered in order to explore the effect of the chromosomes' initial conformation, and the interaction between black chromatin and the lamina, on the nuclear organisation of the chromosomes and proteins. In all cases, 5 independent repeats of each simulation are considered.

Brownian Dynamics simulations are again run using LAMMPS, as described in Chapter 3. The friction due to the solvent is set to $\gamma = m/\tau$, leading to a Brownian timescale $\tau_B = \tau = \tau_{\text{in}}$. The mapping from simulation times to real units could then be done following the approach in Chapter 5, by matching the diffusional properties of the chromatin fibre in *Drosophila* cells [25].

Table 6.1 summarises the simulation parameters.

$\sigma_{\text{chrom}} = \sigma_{\text{prot}} = 30$ nm	TFs
$l_p = 3\sigma = 90$ nm	Chromatin : crowders int. – WCA
chromosomes: 2,3,4 and X	Chromatin : DNA-binding-prot int. – LJ
$M_{\text{Euchr.f}} = 1000$ prot	Enh/Prom : Euchr.f – $\epsilon = 6k_BT$
$M_{\text{HP1}} = 2500$ prot	Transc. genes : Euchr.f – $\epsilon = 3k_BT$
$M_{\text{PRC}} = 2500$ prot	Heterochromatin : HP1 – $\epsilon = 3k_BT$
$M_{\text{crowders}} = 5000$ prot	PcG high affinity sites : PRC – $\epsilon = 6k_BT$
$R_{\text{nucleus}} = 43\sigma = 1.3\mu$m	PcG low affinity sites : PRC – $\epsilon = 3k_BT$
$\gamma_{\text{solvent}} = m/\tau$	$r_{\text{thr}} = 1.8\sigma$
Black chrom. : nuclear wall int. – LJ – $\epsilon = 2k_BT$ and $r_{\text{thr}} = 1.8\sigma$	

Table 6.1 *Simulation parameters for the Drosophila melanogaster genome model.*

6.2 Predicting the formation of chromosome territories

As described above, the system is initialised according to the Rabl orientation, where chromosomes have their centromeres tethered to one point fixed to the nuclear wall, and their ends/telomeres extend to the opposite side of the nucleus (Fig. 6.2(a)). This is achieved by initialising each chromosome arm from a mitotic configuration, i.e. describing a helix of stacked rosettes.

As the system is allowed to evolve and relax, the tight and highly bent rosettes eventually unfold since there is no explicit attraction between polymer beads. Chromosomes thus start adopting a less structured conformation, becoming more ellipsoidal. However, relaxation gets hindered as different chromosomes start to interact and proteins start binding to chromatin. Even though chromosomes slightly mix with each other at the interfaces, overall they remain segregated in distinct territories over the course of a whole simulation, preserving the Rabl orientation (see Fig. 6.3).

This observation is in agreement with previous simulation results [115], where chromosomes initialised as parallel mitotic helices also relax, without mixing, into Rabl-like elongated ellipsoids. The observed chromosome segregation is a consequence of the high level of confinement. Due to the system's high particle

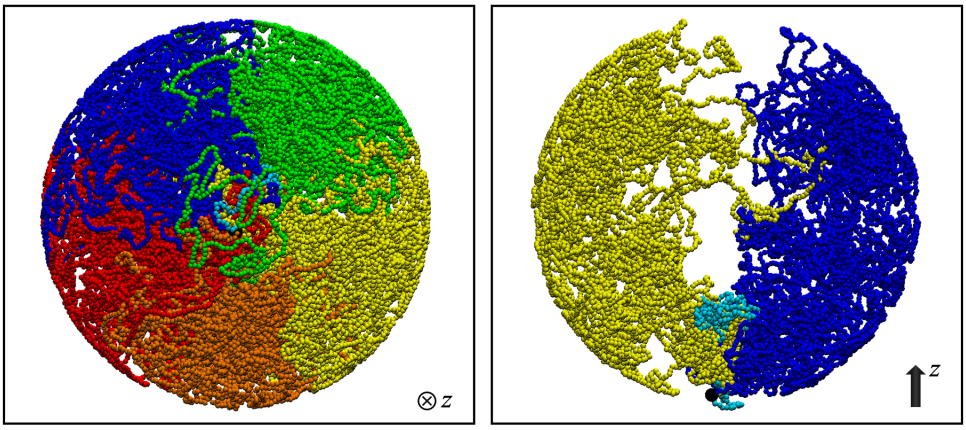

Figure 6.3 *Chromosomes equilibrate, preserving the nuclear organisation into discrete territories. (left) Snapshot showing all chromosomes from the angle of the nuclear tethering point; (right) Snapshot showing chromosome arms 2L and 3L, and chromosome 4. The interaction of black chromatin with the lamina leads to a decrease of chromatin density in the centre of the nucleus.*

volume fraction (8%), the chromosomes behave like polymers in a semi-dilute regime, meaning that very long times are needed for the chromosomes to diffuse enough so as to effectively mix. The results thus suggest that the model correctly captures the formation (or preservation!) of the Rabl-like territorial organisation of chromosomes that is known to persist at interphase in *Drosophila* [80].

Regarding the morphology of the chromosomes, the simulation results are also in agreement with experimental observations from fluorescence in situ hybridisation (FISH) [31, 88, 103] and high-resolution light and electron microscopy [50] studies. *D. melanogaster* chromosomes were observed to be organised during interphase into territories that remain segregated with no significant intermingling, and that span between two opposite poles of the nucleus resembling the segment of an orange [31, 88]. The same conformation is observed in the simulation snapshots in Figure 6.3. Furthermore, the overall size of a chromosome territory has been reported to be roughly determined by its DNA content and other factors such as its transcriptional status [103]. The simulation results are also in agreement with these observations: the size of the simulated territories depends on the length of the chromosomes as well as on the chromosome compaction level, which is regulated by the interaction with proteins/transcription factors. The main difference between the simulation and experimental observations has to do with the choice of simulating an haploid nucleus. Given that in both cases territories span a length of 1 nuclear diameter, in simulations territories span a length of

$\equiv 2.6\mu$m, whereas in experiments these span a length of $\sim 5.2\mu$m [89].

In order to test that the chromosomal nuclear organisation observed in Figure 6.3, which reflects the observed configuration in living cells, simply emerges as a result of the post-mitotic chromosome conformation and the semi-dilute nuclear conditions, another set of simulations was performed where chromosomes were initialised as segregated polymers describing a random walk (RW) trajectory.

To compare the results from both models, HiC-like chromatin interaction maps were calculated according to the procedure previously used in Chapter 5. The maps were generated by recording contacts between chromatin beads whose spatial separation is $< 8\sigma = 240$ nm. Then, these maps were averaged over time (every $10^3\tau$ for the last $10^6\tau$) and over 5 simulation replicates. Figure 6.4 shows the calculated chromatin interaction maps together with the map obtained from HiC experiments [124].

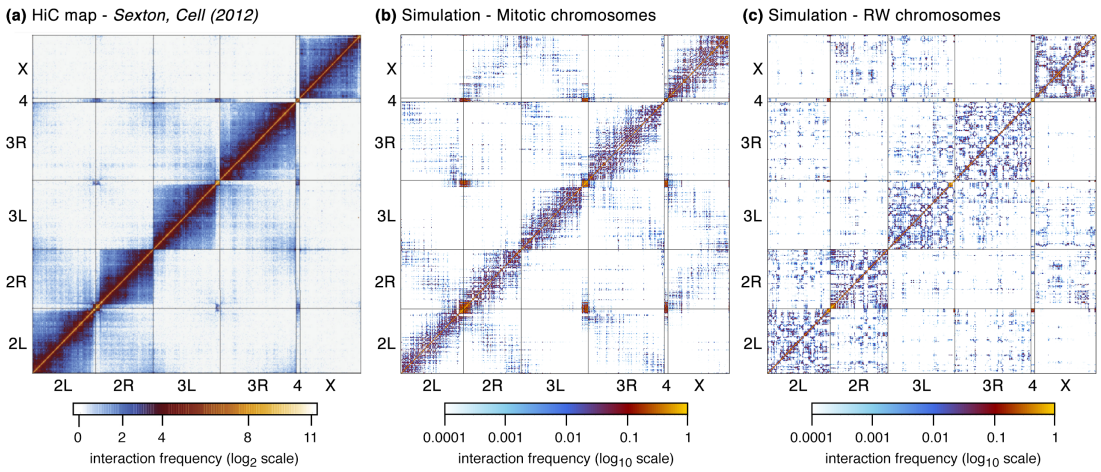

Figure 6.4 *Plots showing chromatin interaction maps for (a) HiC experiments, and simulations where chromosomes are initialised (b) in a mitotic conformation or (c) as random walks. The simulations starting with mitotic chromosomes give the best prediction of the HiC map features, recovering well the interaction signature of the Rabl orientation and the inter-chromosomal interaction patterns.*

Even though both models reproduce the nuclear organisation into distinct chromosome territories, which is reflected in the interaction maps as a strong enhancement of intra over inter-chromosomal interactions, the model starting with mitotic chromosomes indeed gives the best prediction of the overall HiC map features. This clearly captures the interaction signatures of the Rabl orientation: strong interaction domains between centromeric regions, and inter-chromosomal interaction patterns that form a wide stripe perpendicular to the

main map diagonal. The former signature emerges from the fact that all centromeres are tethered to the same point at the nuclear wall, so that those chromosome centromeric regions interact strongly with each other. The latter signature emerges from the fact that the mitotic configuration naturally orients chromosomes parallel to each other so that the telomeres end up grouping at the nuclear pole opposite to the centromeres.

The model starting with RW chromosomes fails to predict these features: the interaction patterns between centromeric regions are much smaller (almost imperceptible), and the inter-chromosomal interaction patterns are "uniformly spread". Therefore, even though randomly initialised chromosomes also relax with little or no mixing, thus preserving the segregation of chromosome territories, the chromatin organisation is such that the interaction frequency between centromeric regions is low and telomeres can even interact frequently with centromeres, as in the case of chromosome arm 2R – see the interaction patterns at the corners of panel 2R in Fig. 6.4(c).

6.3 Predicting the formation of nuclear bodies

At the same time as chromosomes relax and adopt less compact structures, chromatin factors start binding to their cognate chromosome beads. As described in Chapters 2 and 5, proteins naturally cluster due the emerging "bridging-induced attraction", forming clusters of like proteins. Since proteins are here modelled as multivalent spheres, they can bind to chromatin beads irrespective of the chromosome it is embedded in. Therefore, protein clusters can effectively promote inter-chromosomal interactions.

In this work three types of chromatin-binding proteins are considered: euchromatin factors, PcG proteins, and heterochromatin factors (see Sec. 6.1.1.2). Therefore, one observes the formation of three different types of protein clusters (see Fig. 6.7(a)ii.), which bear a huge resemblance to protein nuclear bodies [25, 83].

In a recent experimental study [138], PcG nuclear bodies have been comprehensively characterised in *Drosophila* cells by means of a stochastic optical reconstruction microscopy (STORM) technique. The experimental results indicate the existence of hundreds of PcG nano-clusters, which form through

the polymerization activity of the Polyhomeotic sterile alpha motif (Ph SAM) domain of PRC1. Figure 6.5 shows a STORM image of the observed PcG bodies, along with snapshots of the PcG clusters formed during BD simulations.

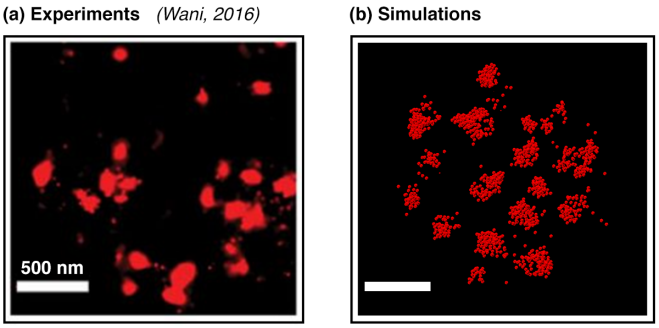

Figure 6.5 *Formation of PcG nuclear bodies. (a) PcG subnuclear bodies imaged by stochastic optical reconstruction microscopy (STORM); (b) Simulation snapshot showing the emerging PcG protein clusters.*

Even though no explicit attraction is set between PcG proteins, the model spontaneously captures the emergence of PcG bodies. More importantly, the model correctly predicts the cluster size distribution of PcG subnuclear clusters observed by STORM imaging (see Fig. 6.6(a)). Since, in the model, PcG clustering is purely driven by PcG protein-chromatin interactions and the bridging-induced attraction mechanism, these results suggest that the chromatin state pattern along the chromosome, together with chromatin and protein concentrations, might be the key factors dictating the size of nuclear bodies. On the other hand, nuclear bodies establish chromatin long-range interactions and even inter-chromosomal interactions. Therefore, chromosomal reorganisation and formation of nuclear bodies are intimately linked after cell division.

From the simulations it was also possible to track all PcG bodies and determine the time dependance of the average PcG cluster diameter. Figure 6.6(b) shows the plot for the mean PcG cluster size, averaged over clusters containing more than 2 proteins, as a function of time.

Three different regimes are observed. For the first $\sim 10\tau_B$, the average cluster size remains practically constant. This corresponds to the time needed for the proteins to diffuse into contact with their chromatin cognate sites. Once proteins start binding to chromatin, PcG clusters also start forming. This regime is marked by an increase of the cluster size with time. According to the Lifshitz–Slyozov theory, which describes the growth of liquid droplets nucleated in vapour, one

should expect the average size to increase with time as $\sim t^{1/3}$ [18]. However, the simulated clusters grow slower, with their average size increasing with time as $\sim t^{0.112}$. This deviation is due to the fact that protein cluster growth is dependent on the dynamics of chromatin [16]. More specifically, protein cluster growth is promoted by the increase in the local concentration of chromatin binding sites, or, in other words, by the condensation of chromatin. However, during the collapse of simple homopolymers, the average polymer cluster/blob size has been shown to increase in time as $\sim t^{0.22}$, which is slower than expected from the Lifshitz–Slyozov theory [21]. In the last regime, protein cluster growth is even slower, indicating that the system starts reaching the point where cluster coarsening becomes arrested.

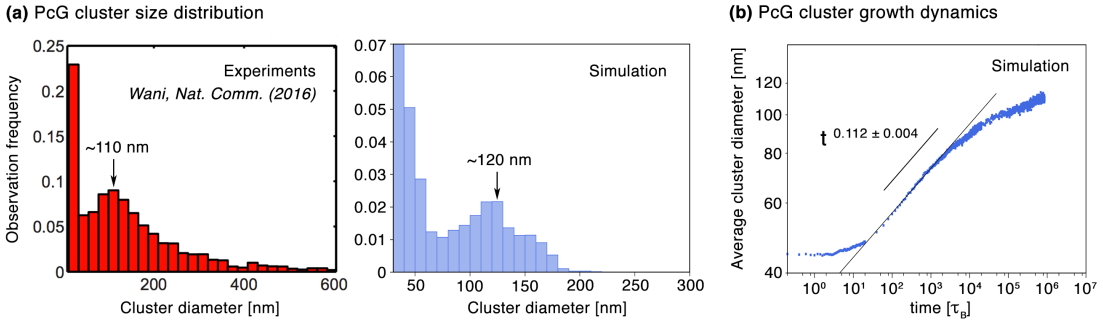

Figure 6.6 *Characterisation of the PcG bodies. (a) Plots showing the PcG cluster size distribution obtained from (left) STORM experiments, and (right) simulations. (b) Plot showing the average PcG cluster diameter as function of time. There are three distinct regimes. In the first regime proteins are diffusing into contact with chromatin. In the second, protein clusters start forming and the average cluster size increases with time as $\sim t^{0.112}$. In the last regime, cluster growth slows down suggesting that cluster coarsening starts becoming arrested.*

6.4 Nuclear distribution of chromatin and nuclear bodies

Given that the model predictions are in good agreement with experimental observations regarding the large-scale organisation of chromosomes and the properties of PcG bodies, the next natural question is how nuclear bodies and different types of chromatin regions (active, inactive, and silenced) are distributed throughout the nucleus.

In order to address these questions, both protein clusters and chromatin beads were tracked and their nuclear radial distribution was recorded. Specifically, the distribution of the locations of the centre of mass of individual protein clusters (with more than two proteins) was recorded. For the different chromatin regions, first each chromatin bead was classified as being active (euchromatin), inactive (heterochromatin or PcG binding sites), or silenced (black chromatin). Then the distribution of the locations of individual beads of each type was recorded. Both distributions (for protein clusters and chromatin) were then averaged over time (every $10^3\tau$ for the last $1.8 \times 10^6\tau$) and over 5 simulation replicates. The final averaged distributions were compared with the distributions obtained for each simulation replica to ensure that the former are not a result of poor statistics: in all cases, the distributions for individual replicas hold the same shape as for the other replicas and the average. Note that, since these are radial distributions, the average number of occurrences for each radial distance is then normalised by the respective spherical surface area. Figure 6.7 summarises the obtained results.

Figure 6.7(a) shows the plot with the radial distribution of the location of protein clusters for euchromatin factors (EuF), PcG proteins (PRC1), and heterochromatin factors (HP1), along with simulation snapshots of the different protein cluster types. Figure 6.7(b) shows the plot with the radial distribution of the location of active, inactive, and silenced chromatin regions, together with simulation snapshots showing the three types of chromatin in red (active), blue (inactive), and grey (silenced).

Figure 6.7(a)i. indicates that HP1 clusters are more or less evenly spread throughout the nucleus. But even so, the distribution reveals a peak towards the nuclear periphery. The simulation snapshots show that, specifically, HP1 clusters tend to aggregate near the chromosome centromeric regions (notice the centromeres' tethering bead in black). This is related to the fact that heterochromatin regions are found prominently near the centromeres and in chromosome 4 [41]. Regarding EuF clusters and PcG (PRC1) bodies, these have a similar radial distribution in the nucleus: both distributions are bimodal with peaks located at the same positions, away from the nuclear centre and periphery. This is also visible from the snapshots on the right, which show that EuF and PcG bodies tend to be close in space. This result agrees with the observations in Ref. [25] which show that PcG bodies never locate within

pericentric heterochromatin[1], but instead tend to be located near more open chromatin regions.

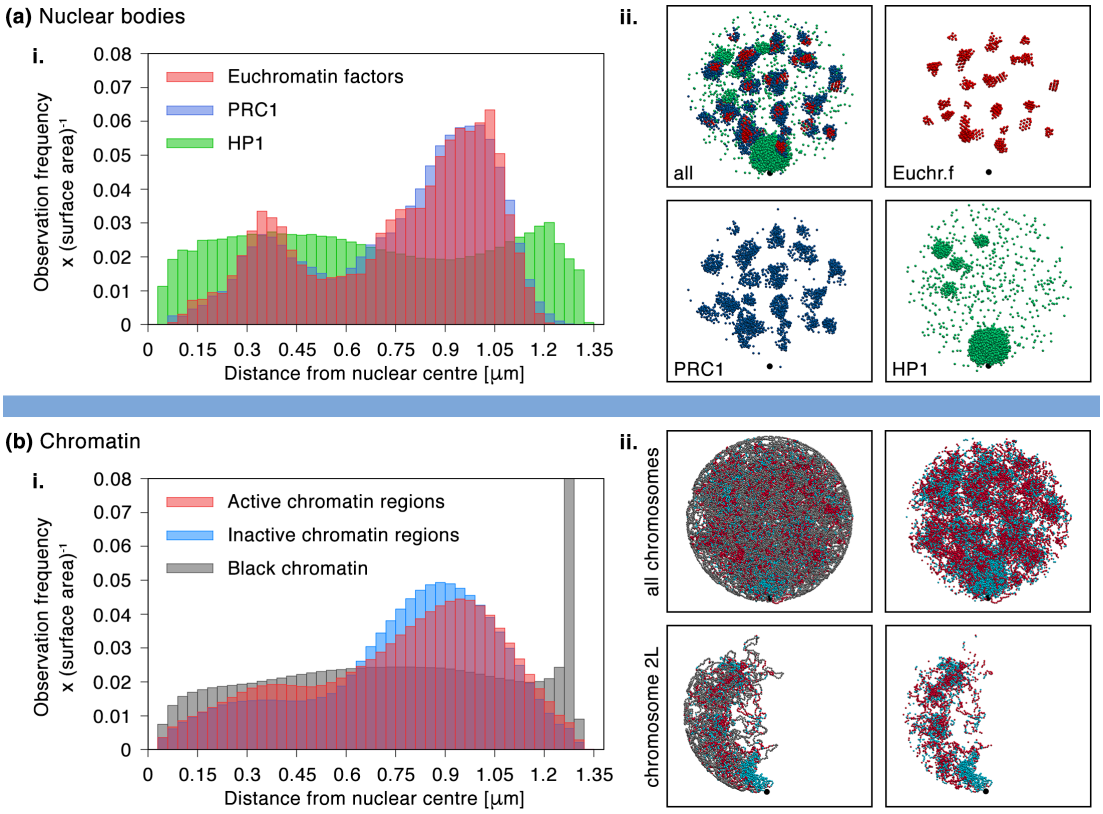

Figure 6.7 *Nuclear distribution of protein clusters and chromatin. (a)i. Plot showing the distribution of the radial position of protein clusters made of euchromatin factors (EuF), PcG proteins (PRC1), and heterochromatin factors (HP1); ii. Simulation snapshots of the different types of nuclear bodies. EuF and PcG bodies are often found close in space, whereas HP1 clusters tend to aggregate near centromeric regions. (b)i. Plot showing the distribution of the radial position of active, inactive, and black chromatin regions; ii. Simulation snapshots showing the three types of chromatin in red (active), blue (inactive), and grey (black). Active and inactive chromatin regions have similar distributions, tending to be positioned away from the nuclear centre. Black chromatin is distributed more evenly, slightly accumulating near the nuclear periphery, due to the interaction with the lamina.*

But even though centromeric regions are highly enriched in heterochromatin, the model yields similar radial distributions for active and inactive chromatin regions (see Fig. 6.7(b)i.). In fact, the distributions for active and inactive chromatin have their peaks located fairly at the same positions as the distributions

[1]Pericentric heterochromatin is a tightly packed form of DNA essentially found at centromeres.

for EuF and PRC1, respectively. This reflects the natural co-localisation of EuFs with euchromatin regions, and PRC1 with PcG affinity regions. On the other hand, black chromatin shows a very distinct distribution: it is more or less uniform across the nucleus, but it shows an intense and narrow peak near the nuclear wall. This is due to the attraction of black chromatin to the nuclear lamina, which, despite being a weak interaction, significantly increases the concentration of silenced regions near the wall.

Interestingly, the observations from both the simulation snapshots and the radial distributions suggest that the concentration of particles in the centre of the nucleus is lower. One explanation for this is that it might come from the chromatin-lamina attraction. Since black chromatin accounts for almost half of the genome, by tending to distribute towards the periphery of the nucleus it might effectively "pull" the chromosomes away from the centre.

6.5 The role of the lamina in nuclear organisation

In order to better assess the effect of the interaction between black chromatin and the nuclear lamina, further simulations were performed where this interaction is turned off. Figure 6.8 summarises the obtained results.

Plot (b) shows the chromatin interaction maps for simulations with lamina associated domains (LADs) on the bottom right, and for simulations without LADs on the top left. These maps suggest a slight increase in the effective concentration of chromatin towards the interior of the nucleus, marked by an enhancement of chromatin interactions: there is an increase of $\sim$10% of the total number of interactions upon the elimination of LADs.

This trend is further supported by the results in plot (c), which show the increase of the average size of PcG clusters with time. In the intermediate growth regime, the average cluster size increases with $t^{0.125}$, i.e. slightly faster than in the case of simulations with LADs, for which the average size grows with $t^{0.112}$. This leads, by the end of the simulation, to the formation of slightly larger protein clusters: the average cluster size, averaged for the last $10^5\tau_B$, increases by $\sim$4.5% when the interaction with the lamina is turned off. I shall stress, though, that this increase of the time exponent and cluster size is very small, meaning that the effect of the lamina here is very subtle. However, since protein cluster growth

is promoted, via the bridging induced attraction, by the increase of the local concentration of chromatin binding sites, together these results support the initial suggestion of an increased effective chromatin nuclear concentration.

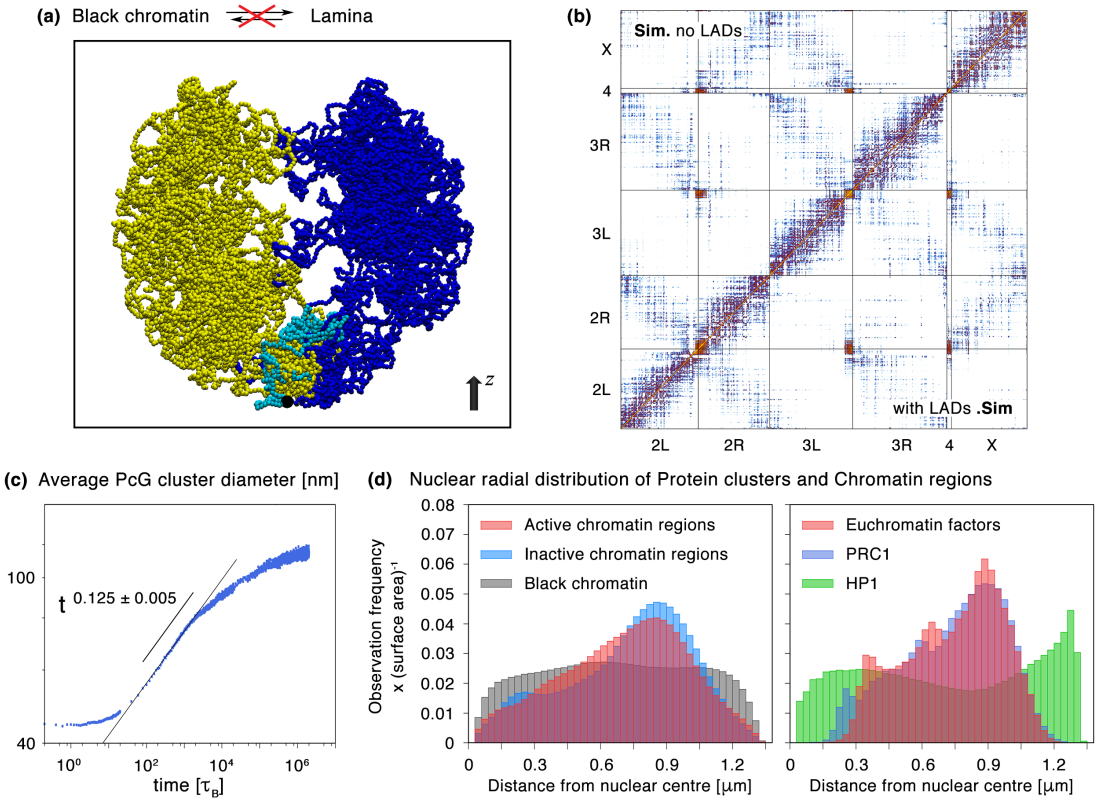

Figure 6.8 *The effect of turning off chromatin-lamina interactions. (a) Simulation snapshot showing chromosome arms 2L, 3L and 4. (b) Plot showing the chromatin interaction maps for simulations with lamina associated domains (LADs) on the bottom right, and without LADs on the top left. The total number of interactions increases by* ∼10% *upon LAD elimination. (c) Plot showing the average PcG cluster diameter as function of time. The average cluster size, averaged for the last* $10^5\tau_B$*, increases by* ∼4.5% *in respect to the case with LADs. (d) Radial distributions for chromatin regions (active, inactive, and silenced) on the left, and for protein clusters (EuF, PRC1, and HP1) on the right.*

The effect of the lamina is more noticeable through the analysis of the nuclear distribution of chromatin and protein clusters (see Fig. 6.8(d)). In the absence of chromatin-lamina interactions, active chromatin regions end up being positioned more towards the centre, and the distribution of inert/black chromatin becomes practically uniform across the whole nucleus. Regarding the radial distributions for euchromatin factors and PcG bodies, these remain fairly similar to each other, but their shape changes substantially in the absence of LADs: they both go from

being bimodal to trimodal distributions.

6.6 Remarks

In summary, in this Chapter the large-scale nuclear organisation of chromatin and proteins in *D. melanogaster* cells was studied by means of Brownian dynamics simulations. By modelling all chromosomes of an haploid cell and associated proteins inside a spherical nucleus, I was able to characterise the restructuring of chromosomal architecture and formation of protein nuclear clusters, after cell division. Previous studies [50, 80, 124, 138] have focused on understanding the properties of the genome and associated proteins during cell interphase. Chromosomes organise into discrete territories, while displaying a polarised nuclear configuration (the Rabl-orientation) [28, 29]: their centromeres cluster at one pole of the nucleus, while their ends/telomeres extend towards the opposite pole, so that chromosomes remain arranged fairly parallel to each other. At the same time, chromatin-bound proteins organise into distinct subnuclear foci and clusters, which, depending on the type of proteins, promote chromatin-chromatin interactions so as to regulate transcription of active genes [102] or repression of inactive genes [138]. The work presented in this chapter aimed at providing further insight on how the interphase organisation is reestablished after disruptive events like cell division.

Therefore, in Section 6.2, I presented simulations of chromosome organisation by three kinds of proteins, where chromosomes were initialised in the mitotic configuration and obeying the polarised Rabl-orientation. I have shown that, as chromosomes relax, they reorganise into less compact ellipsoidal structures, while preserving territorial segregation and the polarised orientation, in agreement with HiC and simulation studies [115, 124]. This suggests that interphase large-scale organisation might be a consequence of the post-mitotic arrangement and the semi-dilute nuclear conditions. In order to test this hypothesis, I then presented another set of simulations where chromosomes were initialised describing random walk (RW) trajectories. Indeed RW chromosomes fail to capture the Rabl-orientation and the long-range intra and inter-chromosome interaction patterns.

In Section 6.3, I have shown that the proposed model also captures the spontaneous formation of protein clusters which largely resemble nuclear bodies,

even though no explicit attraction is set between proteins. Remarkably, the model even correctly predicts the cluster size distribution of Polycomb group (PcG) bodies, measured in recent experiments by means of super-resolution microscopy techniques [138]. In the model, protein cluster formation is driven by protein-chromatin interactions and the bridging-induced attraction mechanism (see Sec 2.3.2), suggesting that the concentration of proteins and chromatin affinity sites might be the key factors regulating the size of nuclear bodies. At the same time, nuclear bodies promote intra and inter-chromosomal interactions, meaning that protein and chromosome nuclear reorganisation are two intimately linked processes.

The nuclear distribution of chromatin and nuclear bodies was analysed in more detail in Section 6.4. I have shown that euchromatin factor (EuF) clusters and PcG bodies tend to be distributed in the nucleus close to each other, with the nuclear radial distributions for both cases revealing two peaks positioned away from the nuclear centre and periphery, whereas heterochromatin factor clusters are more evenly distributed across the nucleus, however tending to aggregate near the chromosome centromeric regions where the density of heterochromatin is high [41]. I have also shown that active and inactive chromatin regions have similar radial distributions, with their peaks matching fairly well the location of the distribution peaks for EuF and PcG bodies, reflecting the co-localisation of nuclear bodies with their chromatin affinity regions. Silenced chromatin regions, however, are more evenly distributed, slightly accumulating towards the periphery due to the interaction with the nuclear lamina. The results seem to further suggest that the concentration of particles towards the nuclear centre is lower. This could be due to the attraction between the nuclear lamina and silenced chromatin regions. Since these account for almost half of the genome, they might effectively "pull" the chromosomes towards the nuclear periphery.

In Section 6.5, I presented another set of simulations where the chromatin-lamina interactions are turned off, so as to better investigate the effect of the lamina in the nuclear organisation. The results indeed suggest that one of the effects of the chromatin-lamina interaction seems to be a slight decrease of the effective chromatin concentration in the interior of the nucleus of haploid cells. This hypothesis is supported by the fact that, in the absence of chromatin-lamina interactions, chromatin interactions are enhanced, and protein cluster growth is faster yielding, by the end of the simulations, slightly larger clusters. Also,

active chromatin regions become distributed more towards the nuclear centre and silenced chromatin regions become evenly distributed across the nucleus. Therefore, it would be of interest to look for this effect in haploid nuclei of *Drosophila* cells in future experiments.

Our results could have functional implications. By keeping silenced chromatin regions "out of the way" near the nuclear wall, the lamina could help lower the effective concentration in the nucleus so as to ease the motion of nuclear components and facilitate DNA repair, DNA replication, and gene regulation processes. It is worth noting that, in this work, the chosen potentials model chromatin-lamina interactions only to a first approximation. Specifically, according to this model, the interaction potential is uniform throughout the nuclear wall, so that any silenced (black) chromatin bead can interact with any wall surface element. It would be, therefore, interesting to explore the effect of a more realistic nuclear lamina model where only specific silenced chromatin loci interact with explicitly modelled membrane proteins.

In summary, the results described in this Chapter suggest that post-mitotic chromosome reorganisation and formation of nuclear bodies are two intimately linked events, where chromatin promotes the nucleation of protein clusters and, in turn, these also promote chromatin-chromatin interactions. The results further suggest that the large-scale interphase chromosomal organisation might emerge mainly as a consequence of the post-mitotic arrangement and semi-dilute nuclear conditions. I shall note that these results are based on a simple model that does not take into account nuclear organelles such as nuclear speckles or the nucleolus, which have been shown to have a significant role in organising chromatin in the nucleus, with active regions being located near nuclear speckles and inactive near the nucleolus [108]. However, since these organelles seem to have a role somewhat similar to the role of the here modelled protein clusters, I would not expect these to have a significant impact on the large scale nuclear organisation of chromosomes.

Conclusions 7

In this Thesis I have made use of a simple coarse-grained polymer model to study the physical properties of DNA in different organisms: from the elastic and dynamic response of bacterial DNA upon compression, to the interphase spatial organisation of human chromosomes, and post-mitotic chromosome and protein organisation in fruit fly nuclei. The remarkable agreement between the simulation results and experimental observations shows how a simple coarse-grained model, which still retains the key elements of each system, can indeed give predictive insights into the properties and behaviour of the genome across different organisms.

In Chapter 4 I presented simulations inspired by recent single molecule experiments, which allow to compress one bacterial chromosome inside a nanopore and measure its elastic and dynamic response. By explicitly modelling the presence of proteins in the system, I was able to quantify their effect on the behaviour of a single DNA molecule. I have shown that proteins which do not bind to DNA, i.e. macromolecular crowders, exert an osmotic pressure on the compression piston that can be orders of magnitude larger than the entropic pressure of the spring-like polymer, dwarfing the DNA entropic response as well as the effect of DNA-binding proteins. Both types of proteins were also shown to significantly hamper the expansion dynamics of DNA after initial compression. The DNA expansion simulations, in the presence of DNA-binding proteins only, further uncovered an interesting protein popping-off kinetics upon tuning the DNA-protein affinity, where proteins are metastably bound to the polymer under confinement but detach one-by-one in solution as DNA starts expanding.

I shall note here that the simulations described in Chapter 4 were performed for polymers torsionally relaxed. However, bacterial DNA is known to be negatively supercoiled *in vivo* (see introduction of Chapter 4). Therefore, an interesting extension to the work presented in this Chapter would be to quantify in detail

the effect of supercoiling on our results.

The organisation of human chromosomes in interphase was studied in Chapter 5. I have conciliated two apparently competing viewpoints regarding the mechanisms responsible for the spatial structure of chromatin: interaction between chromatin and transcription factors (TF) versus chromatin loop extruding complexes (cohesin). Specifically, I have shown that activating and repressive TFs are necessary to fold chromatin into segregated active and inactive domains, and that cohesin complexes are required to organise inert chromatin regions where there is minimal TF binding. I have also presented evidence that cohesin complexes do not necessarily need an active mechanism to organise chromatin. A variation of the TF model was then proposed, yielding a better agreement between simulations and experiments. This suggested that *in vivo* TF binding is not an equilibrium mechanism but rather a dynamic process that allows TFs to bind strongly to chromatin and yet be able to dissociate frequently. The model predictions were further supported by simulations reproducing the effect of recent protein knock-out experiments.

In all simulations performed in Chapter 5, cohesin complexes were considered to be able to bind chromatin irrespective of their genomic location. However, *in vivo* cohesin binding is mediated by the loading factor NIPBL, which binds to chromatin at specific sites. It would be, therefore, interesting to refine our proposed model in future simulations so as to take into account cohesin loaders.

Finally, in Chapter 6, I presented simulations of the nuclear reorganisation of chromosomes and proteins in fruit fly cells, after mitosis (cell division). The results suggest that the large-scale interphase chromosomal organisation is mainly governed by polymer physics principles, being a consequence of the initial mitotic arrangement of chromosomes and the semi-dilute nuclear conditions. I have shown that, as chromosomes relax from mitotic to interphase structures, transcription factors spontaneously organise into clusters resembling nuclear bodies, suggesting that protein and chromosome nuclear organisation are intimately linked through the cell cycle. The results also show that, in haploid fruit fly cells, the distribution of inert chromatin is fairly uniform across the nucleus, whereas active and inactive chromatin tend to be distributed away from the nuclear centre.

In the model proposed in Chapter 6, for chromosome and protein organisation in fruit fly nuclei, folding mechanisms such as chromatin loop extrusion were not

considered. This choice was made on the basis of recent experimental studies which state that this mechanism has a less important effect in fruit flies than in mammals [117]. However, cohesin-like proteins may still have an important role on chromosome architecture, and, thus, it would be of interest to explore the effect of this mechanism in a future simulation study.

All in all, how chromosomes organise in different organisms, so as to fit inside μm-sized cells while keeping expressed genes unfolded and accessible by the cell transcription machinery, will remain a subject of intense study for years to come. This is due to the highly complex cellular composition and interplay between cellular components, which are constantly being uncovered by new experimental techniques. This is a field of great interest to biologists as well as physicists since the understanding of the behaviour of DNA and its interaction with other cell components has major implications in the development of new technological applications and medical treatments. The overall purpose of this thesis was to give my (little) contribution to this vast field of research. I, therefore, aimed at the understanding of some of the mechanisms driving genome organisation in living cells, by making simulation-based predictions which can be tested in, or even inspire, future experiments.

Effect of particle charge on the elasticity of single chromosomes

A

Modelling DNA-protein electrostatic interactions

DNA and proteins are effectively charged in real cells [1]. The phosphates in the DNA backbone are negatively charged, giving DNA molecules an overall negative charge. Specifically, DNA carries two negative charges ($-2e$) per base-pair, due to the negatively charged phosphate groups in each nucleotide. As detailed in Section 3.2.2, 1 DNA bead corresponds to 7.4 bp in the model. Therefore, there are 7.4 phosphate groups per DNA bead. However, counterion (or salt) condensation leads to a neutralisation of $80\% - 100\%$ of the phosphate groups [44]. So, in fact, each DNA bead will only carry a charge of $q_{\mathrm{DNA}} = 0.2 \times (-14.8)e = -2.96e$. The choice of the value of the proteins' charge is not as straight forward since it depends on the protein residues. Here we opted for considering the value of the charge of an average protein, and use that value for negatively and positively charged proteins in the model. In Ref. [119], the average of protein charges in bacteria was measured to lie in the range $[-10e, +15e]$. For simplification we consider the proteins' charge to be equal in magnitude to the DNA beads' charge – $14.8e$ – which corresponds to an effective charge $|q_{\mathrm{prot}}| = 2.96e$ due to neutralisation emerging from counterion condensation. Therefore, crowding proteins are modelled as beads with charge $q = -2.96e$ and DNA-binding proteins with charge $q = +2.96e$.

The electrostatic interactions between charged particles in solution are modelled by considering the Debye-Hückel potential [39, 72], in addition to the

potentials already described in Chapter 3,

$$U_{\mathrm{DH}}(r_{i,i+1}) = C\frac{q_i q_j}{\epsilon_r}\frac{e^{ka}}{1+ka}\frac{e^{-kr_{i,i+1}}}{r_{i,i+1}}, \tag{A.1}$$

where $C = 1/4\pi\epsilon_0 k_B T$, q_i is the charge of particle i, ϵ_r the dimensionless dielectric constant (we consider $\epsilon_r = \epsilon_{r,\ \mathrm{water}} = 80$), k the inverse Debye length, and a the radius of the particle.

We consider $k^{-1} = 1$ nm, which is the Debye length in the cell medium. More explicitly, $k = \sqrt{8\pi l_B N_A 10^3 c_S}$, where $l_B = 0.71$ nm is the Bjerrum length in water and $c_S \sim 150$ mM is the salt concentration inside cells [3].

Comparing the effect of particle charge

In this Section I present the results corresponding to Figures 4.4, 4.5 and 4.6 in Chapter 4, but for charged DNA beads and proteins.

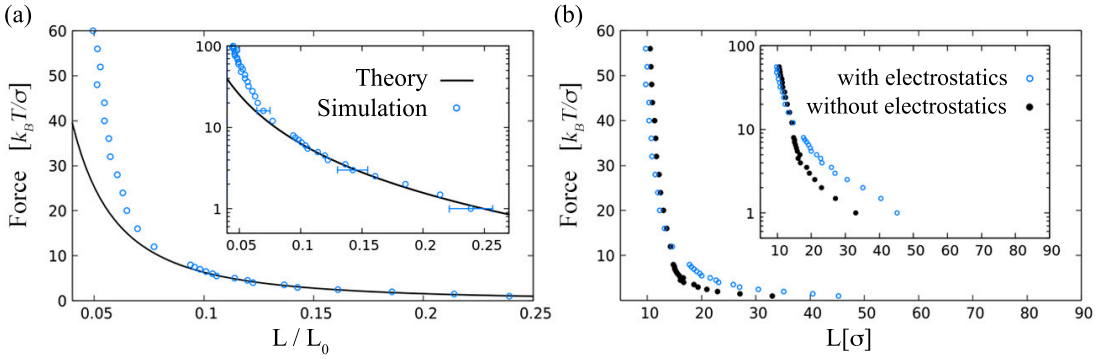

Figure A.1 *As Figure 4.4, but for simulations with charged polymer beads. (a) Comparison of the force-extension curve obtained from the entropic spring theory of Ref. [104] (full line) and the numerical model for charged particles (blue circles). Like in the simulations without charge, the entropic spring theory of Ref. [104] agrees with the model for $L/L_0 > 0.1$, but breaks down for $L/L_0 < 0.1$. (b) Comparison of the force-extension curves obtained from the model with (blue circles) and without (black dots) electrostatic interactions. For $f > 10$, both models lead to similar force-extension relations. For $f < 10$, the extension of the charged polymer for a given force is larger than for the neutral case, which is expected since there is an additional longer-range repulsion between DNA beads.*

Figure A.1 shows the force extension curves in the absence of proteins. In plot (a), like in Figure 4.4, the scaling theory of a linear string of polymer blobs agrees with the model for $L/L_0 > 0.1$, but breaks down for $L/L_0 < 0.1$, where excluded

volume effects become significant. Figure A.1(b) shows that, for $f > 10$, the charge of DNA beads does not play a significant role in the elastic response of the polymer, but that for weak compression forces ($f < 10$) the effect of long-range electrostatic interactions is more noticeable: for a given force the DNA extension for the charged polymer is larger than for the neutral polymer.

Figure A.2 for charged particles and Figure 4.5 for neutral particles are remarkably similar. The comparison between inset 2 in both Figures shows, however, that the presence of electrostatic interactions leads to a slightly higher DNA extension for a moderate force, as seen before.

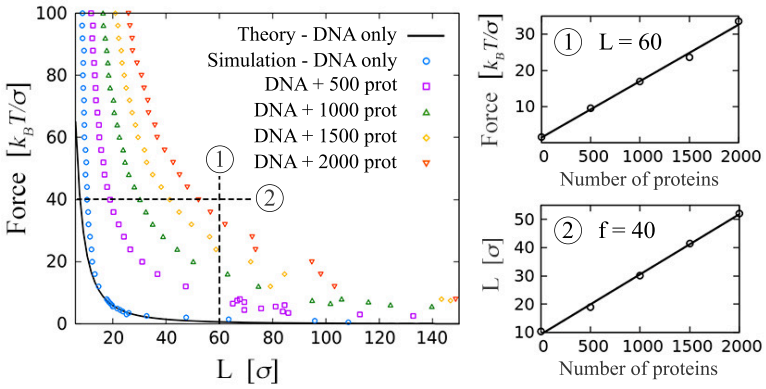

Figure A.2 *As Figure 4.5, but for simulations with charged proteins and DNA beads – Force-extension curves for a varying number M of non-DNA-binding proteins. The force-extension curves are remarkably similar to the ones in Figure 4.5. The linear relations force - M (inset 1 – fit: $f = 0.0156M + 1.4513$), and DNA extension - M (inset 2 – fit: $f = 0.0212M - 9.3741$) are also recovered. The comparison between inset 2 and the one in Figure 4.5 shows that the presence of electrostatic interactions leads to a slightly higher DNA extension for a moderate force, as seen in Figure A.1.*

Again, Figure A.3 and Figure 4.6 lead to the same conclusions. In both cases the DNA-binding proteins lead to the formation of DNA clusters, hence compacting the DNA, giving a decrease in the compression force. The striking similarity between the results arising from the charged and neutral models further suggests that the electrostatic interactions do not play a significant role in the overall elastic response of DNA.

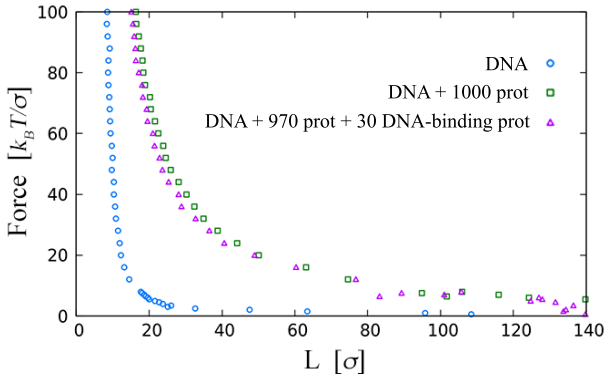

Figure A.3 *As Figure 4.6, but for simulations with charged proteins and polymer beads – Comparison of the force-extension curves for simulations with just DNA (blue circles), DNA and crowding proteins (green squares), and DNA in the presence of both binding and non-binding proteins (magenta triangles: 3% of the proteins are DNA-binding). Like in the simulations presented in Figure 4.6, DNA-binding proteins lead to the formation of DNA clusters, hence compacting the DNA, giving a decrease in the compression force.*

Effect of protein size on the elasticity and dynamics of single chromosomes

B

In Chapter 4 the effect of proteins on the elasticity and dynamics of bacterial DNA was investigated considering proteins modelled as spherical beads of the size of the DNA polymer beads: $\sigma \equiv 2.5$ nm. This was a choice done on the basis of simplicity. However, bacterial proteins in real cells have an average size of ~ 5 nm [106], which corresponds to twice the thickness of B-DNA. Therefore, in the model proposed in this Thesis, one would consider more realistic proteins to have a size of 2σ. In this Appendix, the effect of larger proteins (size - 2σ) on the DNA behaviour is explored, focusing on the comparison with the results of Figures 4.5, 4.6 and 4.7 in Chapter 4.

Figure B.1 shows the force-compression curves for simulations with varying number of crowding proteins: comparison between these results and those of Figure 4.5 show a very similar trend. The quantitative values are very close together for large enough L, and start to deviate significantly below $L \sim 60 - 70$, where the volume fraction is non-negligible, and excluded volume effects are larger for the larger proteins.

The comparison between Figure B.2 and Figure 4.6 leads to the same conclusion – the slightly larger gap between cases with and without DNA-binding proteins is due to the fact that larger DNA-binding proteins lead to multiple binding to the DNA, hence the DNA becomes more compact. Thus the observed effect might not just be due to protein size, as bacterial DNA-binding proteins such as H-NS only have two DNA-binding sites (they are better represented in our model by the smaller proteins considered in Chapter 4).

The effect of the larger protein size on the expansion dynamics is analysed in

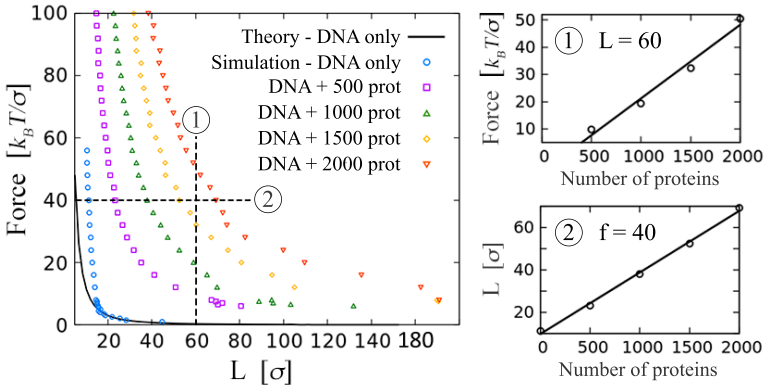

Figure B.1 *As Figure 4.5, but for simulations with proteins of diameter* 2σ*. – Compression force as a function of the DNA extension for a varying number* M *of non-DNA-binding proteins. As for the simulations presented in Figure 4.5, the protein osmotic contribution leads to a large increase in the compression force, sometimes of several orders of magnitude. For a given value of* L*, the protein contribution is linear in the number of proteins (inset 1) – fit:* $f = 0.0271M - 5.8389$*. For a given compression force, the DNA extension also increases linearly with the increasing number of proteins (inset 2) – fit:* $f = 0.0290M + 9.8640$*.*

Figure B.3, which shows a slower expansion dynamics. This is because larger spheres create larger effective friction (as the friction, or viscosity of a hard sphere suspension is proportional to its volume fraction). The measured effective exponent is accordingly smaller, although this should be seen as a measure of the speed of the dynamics rather than a true dynamical exponent, which would require the study of different chain lengths. The decrease in the effective exponent is more marked in the case with DNA-binding proteins (although the change in the exponent, with respect to the case just with crowding proteins, is small or even not significant within the exponents' error bars): again, this is due to the fact that these proteins can form multiple contacts with DNA.

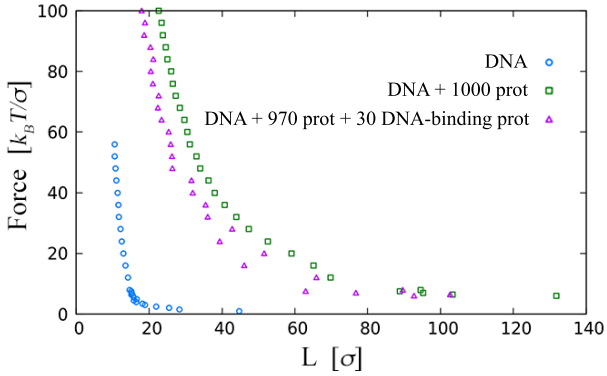

Figure B.2 *As Figure 4.6, but for simulations with proteins of diameter* 2σ*. – Compression force as a function of the DNA extension for simulations with just DNA (blue circles), DNA and crowding proteins proteins (green squares), and DNA in the presence of both binding and non-binding proteins (magenta triangles:* 3% *of the proteins are DNA-binding). Like in the simulations presented in Figure 4.6, the DNA-binding proteins lead to the formation of DNA clusters, hence compacting the DNA, giving a decrease in the compression force. However, since here proteins are larger, there is multiple binding to DNA, which leads to a more compact DNA. This explains the slightly larger gap between the curves for the cases with and without DNA-binding proteins.*

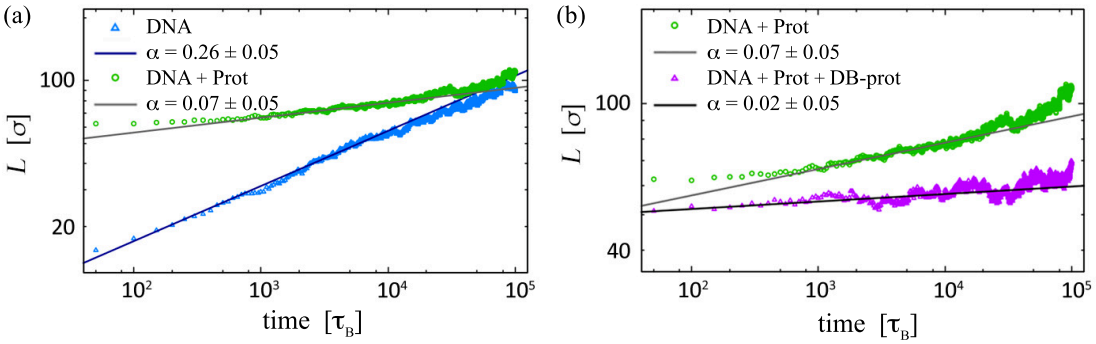

Figure B.3 *As Figure 4.7, but for simulations with proteins of diameter* 2σ*. – Measured dynamical exponents during DNA expansion (a) for simulations without proteins (blue triangles) and with* 1000 *crowding proteins (green circles). Like in the simulations presented in Figure 4.7, the protein crowding leads to a lower exponent, however here the exponent decrease is larger; and (b) for simulations in the presence of* 970 *crowding and* 30 *DNA-binding proteins (magenta triangles). The exponent decreases in the presence of DNA-binding proteins, like in the case of Figure 4.7, however here large proteins lead to a more marked decrease of the exponent. This is due to the fact that larger proteins allow multiple binding to DNA, which further decreases the DNA extensional elasticity. All simulations started from an initial DNA configuration obtained for a compression force of* 20*. The range of times, for which the expansion curves were fit, were chosen to take into account the intermediate expansion regime (*$10^3\tau_B$ – $10^4\tau_B$ *or* 0.01 *ms* – 0.1 *ms).*

Power-law behaviour of the DNA extension during free expansion

C

In Chapter 4 the expansion dynamics, after initial compression, of a single bacterial chromosome trapped in a cylindrical pore was studied in the absence and presence of crowding and DNA-binding proteins. The same study was performed in Appendix B to explore the effect of the protein size on the DNA behaviour. The polymer expansion dynamics was analysed by measuring the dynamical exponent α, assuming that the polymer extension along the pore, L, increases with time in a power-law fashion – $L(t) \sim t^{\alpha}$ (Eq. (2.27)) – in the intermediate expansion regime after the initial abrupt expansion, as expected from the entropic spring theory for self-avoiding polymers revised in Chapter 2.

In both chapters, the simulations were performed considering a DNA segment modelled as a polymer of $N = 1000$ beads, which corresponds to a chromosome segment of $\sim 2.5\mu$m. In this Appendix the expansion dynamics of a longer DNA polymer, with $N = 2000$ beads corresponding to a length of $\sim 5\mu$m, is studied in order to ascertain the validity of the assumption of the power-law behaviour $L(t) \sim t^{\alpha}$, focusing on the comparison with the results of Figures 4.7 and B.3, in Chapter 4 and Appendix B respectively.

For that, the DNA expansion simulations reported in Section 4.3 and Appendix B were repeated, but this time for a 2000 bead-long polymer. All other parameters were left unchanged, including the number of protein crowders ($M = 1000$ or $M = 970$) and DNA-binding proteins ($M = 30$). Figure C.1 summarises the expansion dynamics results for the 2000 bead-long polymer, for the cases (a) $\sigma_{\text{prot}} = \sigma_{\text{DNA}} \equiv 2.5$ nm and (b) $\sigma_{\text{prot}} = 2\sigma_{\text{DNA}} \equiv 5$ nm, and Table C.1 compares the dynamical exponents, α, measured for the intermediate expansion

regime with those in Figures 4.7 and B.3. Overall the simulations for a longer DNA polymer yield the same values for α, given the error in the exponents, corroborating the assumption that the DNA extension along the cylindrical pore indeed increases with time as $L(t) \sim t^{\alpha}$ after the initial abrupt expansion.

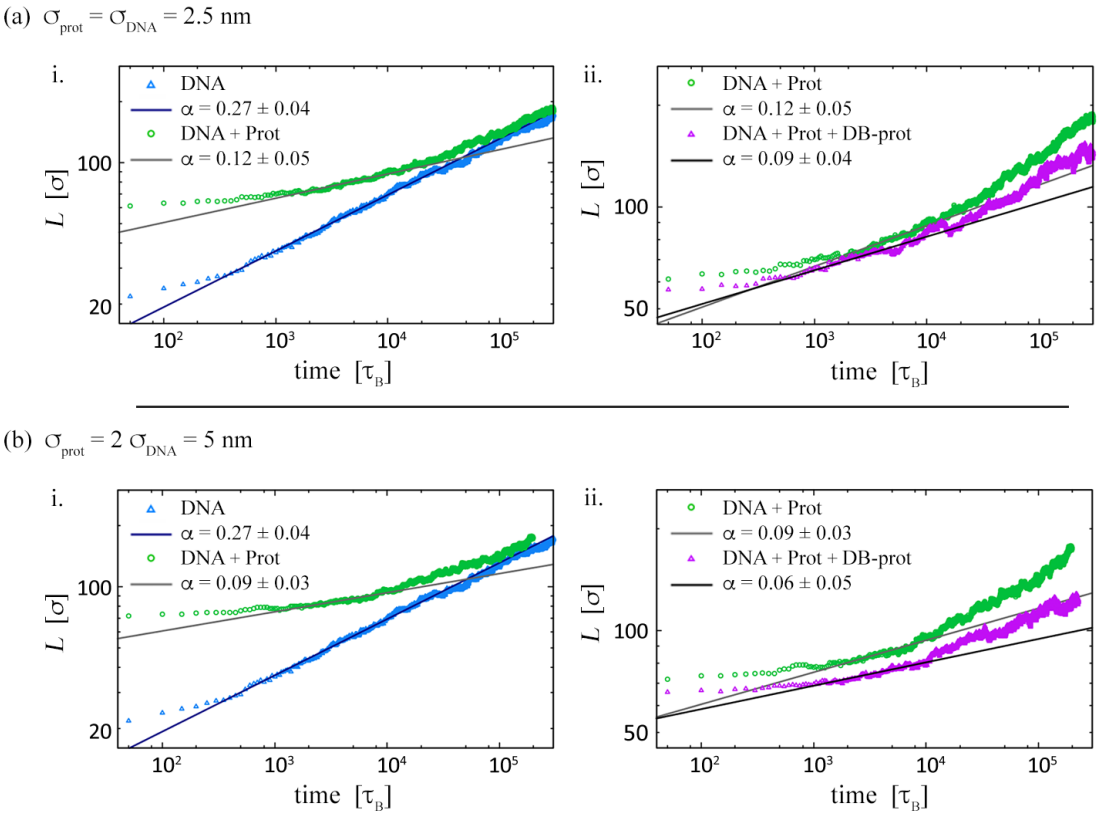

Figure C.1 *As (a) Figure 4.7 and (b) Figure B.3, but for simulations with a* 2000 *bead-long DNA polymer. – Measured dynamical exponents during DNA expansion (i) for simulations without proteins (blue triangles) and with* 1000 *crowding proteins (green circles); and (ii) for simulations in the presence of* 970 *crowding and* 30 *DNA-binding proteins (magenta triangles). All simulations started from an initial DNA configuration obtained for a compression force of* 20*. The range of times, for which the expansion curves were fit, were chosen to take into account the intermediate expansion regime* ($10^3\tau_B$ – $10^4\tau_B$ *or* 0.01 *ms* – 0.1 *ms).*

Measured dynamical exponents		
Polymer length	$N = 1000$ beads	$N = 2000$ beads
$\sigma_{\text{prot}} = \sigma_{\text{DNA}}$	$\alpha_{\text{DNA}} = 0.26 \pm 0.05$ $\alpha_{\text{DNA+Prot}} = 0.10 \pm 0.05$ $\alpha_{\text{DNA+DB-Prot}} = 0.07 \pm 0.05$	$\alpha_{\text{DNA}} = 0.27 \pm 0.04$ $\alpha_{\text{DNA+Prot}} = 0.12 \pm 0.05$ $\alpha_{\text{DNA+DB-Prot}} = 0.09 \pm 0.04$
$\sigma_{\text{prot}} = 2\sigma_{\text{DNA}}$	$\alpha_{\text{DNA}} = 0.26 \pm 0.05$ $\alpha_{\text{DNA+Prot}} = 0.07 \pm 0.05$ $\alpha_{\text{DNA+DB-Prot}} = 0.02 \pm 0.05$	$\alpha_{\text{DNA}} = 0.27 \pm 0.04$ $\alpha_{\text{DNA+Prot}} = 0.09 \pm 0.03$ $\alpha_{\text{DNA+DB-Prot}} = 0.06 \pm 0.05$

Table C.1 *Comparison of the measured dynamical exponents, α, during DNA expansion for simulations with a polymer of (left)* 1000 *beads, and (right)* 2000 *beads. Given the error in the exponents, the measured values for α are the same for the two simulated polymer lengths, supporting the assumption that the DNA extension along the cylindrical pore indeed shows the power-law behaviour $L(t) \sim t^{\alpha}$, during the intermediate expansion regime.*

Bibliography

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 2002.

[2] E. Alipour and J. F. Marko. Self-organization of domain structures by dna-loop-extruding enzymes. *Nucleic Acids Research*, 40:11202, 2012.

[3] T. Ando and J. Skolnick. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc. Natl. Acad. Sci. USA*, 107:18457–18462, 2010.

[4] A. Angel, J. Song, C. Dean, and M. Howard. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature*, 476:105–108, 2011.

[5] S. Asakura and F. Oosawa. Interactions between particles suspended in solutions of macromolecules. *J. Polym. Sci. [B]*, 33:183–192, 1958.

[6] F. Aumann, F. Lankas, M. Caudron, and J. Langowski. Monte carlo simulation of chromatin stretching. *Phys. Rev. E*, 73:041927, 2006.

[7] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences USA*, 109:16173–16178, 2012.

[8] A. D. Bates and A. Maxwell. *DNA topology*. Oxford University Press, New York, 2005.

[9] R. A. Beagrie, A. Scialdone, M. Schueler, D. C. A. Kraemer, M. Chotalia, S. Q. Xie, M. Barbieri, I. de Santiago, L.-M. Lavitas, M. R. Branco, J. Fraser, J. Dostie, L. Game, N. Dillon, P. A. W. Edwards, M. Nicodemi, and A. Pombo. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, 543:519–524, 2017.

[10] V. G. Benza, B. Bassetti, K. D. Dorfman, V. F. Scolari, K. Bromek, P. Cicuta, and M. C. Lagomarsino. Physical descriptions of the bacterial nucleoid at large scales, and their biological implications. *Rep. Progr. Phys.*, 75:076602, 2012.

[11] S. Bianco, D. G. Lupiáñez, A. M. Chiariello, C. Annunziatella, K. Kraft, R. Schöpflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, S. Mundlos,

and M. Nicodemi. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nature Genetics*, 50:662–667, 2018.

[12] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, and V. Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophysical Journal*, 76:409–413, 1999.

[13] C. A. Brackley, J. M. Brown, D. Waithe, C. Babbs, J. Davies, J. R. Hughes, V. J. Buckle, and D. Marenduzzo. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biology*, 17:59, 2016.

[14] C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook, and D. Marenduzzo. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucl. Acids Res.*, 44:3503–3512, 2016.

[15] C. A. Brackley, J. Johnson, D. Michieletto, A. N. Morozov, M. Nicodemi, P. R. Cook, and D. Marenduzzo. Non-equilibrium chromosome looping via molecular slip-links. *Physical Review Letters*, 119:138101, 2017.

[16] C. A. Brackley, B. Liebchen, D. Michieletto, F. L. Mouvet, P. R. Cook, and D. Marenduzzo. Ephemeral protein binding to dna shapes stable nuclear bodies and chromatin domains. *Biophysical Journal*, 28:1085, 2017.

[17] C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, and D. Marenduzzo. Nonspecific bridging-induced attraction drives clustering of dna-binding proteins and genome organization. *Proc. Natl. Acad. Sci. USA*, 110:E3605–E3611, 2013.

[18] A. J. Bray. Theory of phase-ordering kinetics. *Advances in Physics*, 51:481–587, 2002.

[19] D. F. Browning, D. C. Grainger, and S. J. W. Busby. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.*, 13:773–780, 2010.

[20] G. A. Busslinger, R. R. Stocsits, P. van der Lelij, E. Axelsson, A. Tedeschi, N. Galjart, and J.-M. Peters. Cohesin is positioned in mammalian genomes by transcription, ctcf and wapl. *Nature*, 544:503–507, 2017.

[21] A. Byrne, P. Kiernan, D. Green, and K. A. Dawson. Kinetics of homopolymer collapse. *J. Chem. Phys.*, 102:573–577, 1995.

[22] C. R. Calladine, H. R. Drew, B. F. Luisi, and A. A. Travers. *Understanding DNA: The Molecule and How It Works*. Elsevier Academic Press, 2004.

[23] G. Cavalli and T. Misteli. Functional implications of genome topology. *Nat. Struct. Mol. Biol.*, 20:290–299, 2013.

[24] J. Chen, Z. Zhang, L. Li, B. C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, and Z. Liu. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, 156:1274–1285, 2014.

[25] T. Cheutin and G. Cavalli. Progressive polycomb assembly on h3k27me3 compartments generates polycomb bodies with developmentally regulated motion. *PLoS Genetics*, 8:e1002465, 2012.

[26] S. Chong, C. Dugast-Darzacq, Z. Liu, P. Dong, G. Dailey, C. Cattoglio, A. Heckert, S. Banala, L. Lavis, X. Darzacq, and R. Tjian. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, 361:eaar2555, 2018.

[27] P. R. Cook. The Organization of Replication and Transcription. *Science*, 284:1790–1795, 1999.

[28] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, 2:292–301, 2001.

[29] T. Cremer and C. Cremer. Rise, fall and resurrection of chromosome territories: a historical perspective. part i. the rise of chromosome territories. *European Journal of Histochemestry*, 50:161–176, 2006.

[30] S. W. Criscione, Y. V. Teo, and N. Neretti. The chromatin landscape of cellular senescence. *Trends in Genetics*, 32:751–761, 2016.

[31] A. K. Csink and S. Henikoff. Large-scale chromosomal movements during interphase progression in *Drosophila*. *The Journal of Cell Biology*, 143:13––22, 1998.

[32] R. T. Dame, M. C. Noom, and G. J. L. Wuite. Bacterial chromatin organization by h-ns protein unravelled using dual dna manipulation. *Nature*, 444:387–390, 2006.

[33] I. F. Davidson, D. Goetz, M. P. Zaczek, M. I. Molodtsov, P. J. Huis in 't Veld, F. Weissmann, G. Litos, D. A. Cisneros, M. Ocampo-Hafalla, R. Ladurner, F. Uhlmann, A. Vaziri, and J.-M. Peters. Rapid movement and transcriptional re-localization of human cohesin on dna. *The EMBO Journal*, 35:2671–2685, 2016.

[34] P. G. de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.

[35] R. de Vries. Dna compaction by nonbinding macromolecules. *Polym. Sci. Ser. C*, 54:30–35, 2012.

[36] S. C. Dillon and C. J. Dorman. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, 8:185–195, 2010.

[37] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485:376–380, 2012.

[38] M. Doi and S. F. Edward. *The Theory of Polymer Dynamics*. Oxford University Press, New York, 1986.

[39] J. D. Dwyer and V. A. Bloomfield. Brownian dynamics simulation of probe diffusion in dna: effects of probe size, charge and dna concentration. *Biophys. Chem.*, 57:55–64, 1995.

[40] ENCODE Project Consortium. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489:57–74, 2012.

[41] G. J. Filion, J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, and B. van Steensel. Systematic protein location mapping reveals five principal chromatin types in drosophila cells. *Cell*, 143:212–224, 2010.

[42] P. J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, 1953.

[43] N. E. Follmer, A. H. Wani, and N. J. Francis. A polycomb group protein is retained at specific sites on chromatin in mitosis. *PLoS Genetics*, 8:e1003135, 2012.

[44] C. Forrey and M. Muthukumar. Langevin dynamics simulations of genome packing in bacteriophage. *Biophys. J.*, 91:25–41, 2006.

[45] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny. Formation of chromosomal domains by loop extrusion. *Cell Reports*, 15:2038–2049, 2016.

[46] G. L. Hager, J. G. McNally, and T. Misteli. Transcription dynamics. *Molecular Cell*, 35:741–753, 2009.

[47] P. J. Hagerman. Flexibility of dna. *Annu. Rev. Biophys. Biophys. Chem.*, 17:265–286, 1988.

[48] H. Hajjoul, J. Mathon, H. Ranchon, I. Goiffon, J. Mozziconacci, B. Albert, P. Carrivain, Victor J.-M., O. Gadal, K. Bystricky, and A. Bancaud. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res.*, 23:1829–1838, 2013.

[49] A. S. Hansen, C. Cattoglio, X. Darzacq, and R. Tjian. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*, 1034:1–13, 2017.

[50] M. Hochstrasser, D. Mathog, Y. Gruenbaum, H. Saumweber, and J. W. Sedat. Spatial organization of chromosomes in the salivary gland nuclei of *Drosophila melanogaster*. *The Journal of Cell Biology*, 102:112–123, 1986.

[51] J. R. Hughes, N. Roberts, S. McGowan, D. Hay, E. Giannoulatou, M. Lynch, M. De Gobbi, S. Taylor, R. Gibbons, and D. R. Higgs. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*, 46:205–212, 2014.

[52] T. Hunter and M. Karin. The regulation of transcription by phosphorylation. *Cell*, 70:375–387, 1992.

[53] B. V. S. Iyer, M. Kenward, and G. Arya. Hierarchies in eukaryotic genome organization: Insights from polymer theory and simulations. *BMC Biophys.*, 4:8, 2011.

[54] J. Johnson, C. A. Brackley, P. Cook, and D. Marenduzzo. A simple model for dna bridging proteins and bacterial or human genomes: bridging-induced attraction and genome compaction. *Journal of Physics: Condensed Matter*, 27:064119, 2015.

[55] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, 42:9553–9561, 2014.

[56] M. Joyeux. Compaction of bacterial genomic dna: clarifying the concepts. *J. Phys. Cond. Matt.*, 27:383001, 2015.

[57] S. Jun, A. Arnold, and B.-Y. Ha. Confined space and effective interactions of multiple self-avoiding chains. *Phys. Rev. Lett.*, 98:128303, 2007.

[58] S. Jun and B. Mulder. Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proc. Natl. Acad. Sci. USA*, 103:12388–12393, 2006.

[59] S. Jun, D. Thirumalai, and B.-Y. Ha. Compression and stretching of a self-avoiding chain in cylindrical nanopores. *Phys. Rev. Lett.*, 101:138101, 2008.

[60] S. Jun and A. Wright. Entropy as the driver of chromosome segregation. *Nat. Rev. Microbiol.*, 8:600–607, 2010.

[61] Y. Jung, C. Jeon, M. Ha, and B.-Y. Ha. Expansion dynamics of a self-avoiding polymer in a cylindrical pore. *EPL*, 104:68003, 2013.

[62] Y. Jung, C. Jeon, J. Kim, H. Jeong, S. Jun, and B.-Y. Ha. Ring polymers as model bacterial chromosomes: Confinement, chain topology, single chain statistics, and how they interact. *Soft Matter*, 8:2095–2102, 2012.

[63] Y. Jung, S. Jun, and B.-Y. Ha. Self-avoiding polymer trapped inside a cylindrical pore: Flory free energy and unexpected dynamics. *Phys. Rev. E*, 79:061912, 2009.

[64] I. Junier, F. Boccard, and O. Espéli. Polymer modeling of the e. coli genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation. *Nucl. Acids Res.*, 42:1461–1473, 2014.

[65] M. Kanke, E. Tahara, P. J. Huis in't Veld, and T. Nishiyama. Cohesin acetylation and wapl-pds5 oppositely regulate translocation of cohesin along dna. *The EMBO Journal*, 35:2686–2698, 2016.

[66] P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, D. Linder-Basso, A. Plachetka, G. Shanower, M. Y. Tolstorukov, L. J. Luquette, R. Xi, Y. L. Jung, R. W. Park, E. P. Bishop, T. K. Canfield, R. Sandstrom, R. E. Thurman, D. M. MacAlpine, J. A. Stamatoyannopoulos, M. Kellis, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, G. H. Karpen, and P. J. Park. Comprehensive analysis of the chromatin landscape in drosophila melanogaster. *Nature*, 471:480–486, 2011.

[67] A. Khorshid, S. Amin, Z. Zhang, T. Sakaue, and W. W. Reisner. Nonequilibrium dynamics of nanochannel confined dna. *Macromolecules*, 49:1933, 2016.

[68] A. Khorshid, P. Zimny, D. T.-La Roche, G. Massarelli, T. Sakaue, and W. Reisner. Dynamic compression of single nanochannel confined dna via a nanodozer assay. *Phys. Rev. Lett.*, 113:268104, 2014.

[69] J. Kim, C. Jeon, H. Jeong, Y. Jung, and B.-Y. Ha. A polymer in a crowded and confined space: effects of crowder size and poly-dispersity. *Soft Matter*, 11:1877–1888, 2015.

[70] H. Kimura, K. Sugaya, and P. R. Cook. The transcription cycle of rna polymerase ii in living cells. *J Cell Biol*, 159:777–782, 2002.

[71] N. Kubo, H. Ishii, D. Gorkin, F. Meitinger, X. Xiong, R. Fang, T. Liu, Z. Ye, B. Li, J. Dixon, A. Desai, H. Zhao, and B. Ren. Preservation of chromatin organization after acute loss of ctcf in mouse embryonic stem cells. *bioRxiv*, 118737, 2017.

[72] K. K. Kunze and R. R. Netz. Salt-induced dna-histone complexation. *Phys. Rev. Lett.*, 85:4389–4392, 2000.

[73] J. Langowski. Polymer chain models of dna and chromatin. *The European Physical Journal E*, 19:241–249, 2006.

[74] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar. Liquid droplet formation by hp1α suggests a role for phase separation in heterochromatin. *Nature*, 547:236–240, 2017.

[75] T. B. K. Le, M. V. Imakaev, L. A. Mirny, and M. T. Laub. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, 342:731–734, 2013.

[76] J. C. Le Guillou and J. Zinn-Justin. Critical exponents for the n-vector model in three dimensions from field theory. *Phys. Rev. Lett.*, 39:95–98, 1977.

[77] X. Li, C. M. Schroeder, and K. D. Dorfman. Modeling the stretching of wormlike chains in the presence of excluded volume. *Soft Matter*, 11:5947–5954, 2015.

[78] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326:289–293, 2009.

[79] Z. Liu, W. R. Legant, B. C. Chen, L. Li, J. B. Grimm, L. D. Lavis, E. Betzig, and R. Tjian. 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *Elife*, 3:e04236, 2014.

[80] M. G. Lowenstein, T. D. Goddard, and J. W. Sedat. Long-range interphase chromosome organization in drosophila: A study using color barcoded fluorescence in situ hybridization and structural clustering analysis. *Molecular Biology of the Cell*, 15:5678–5692, 2004.

[81] D. G. Lupiáñez, M. Spielmann, and S. Mundlos. Breaking tads: How alterations of chromatin domains result in disease. *Trends in Genetics*, 32:225–237, 2016.

[82] Y. Mao, M. E. Cates, and H. N. W. Lekkerkerker. Depletion force in colloidal systems. *Physica A*, 222:10–24, 1995.

[83] Y. S. Mao, B. Zhang, and D. L Spector. Biogenesis and function of nuclear bodies. *Trends in Genetics*, 27:295–306, 2011.

[84] A. Marantan and L. Mahadevan. Mechanics and statistics of the worm-like chain. *American Journal of Physics*, 86:86–94, 2018.

[85] D. Marenduzzo, K. Finan, and P. R. Cook. The depletion attraction: an underappreciated force driving cellular organisation. *J. Cell Biol.*, 175:681–686, 2006.

[86] D. Marenduzzo, E. Orlandini, and C. Micheletti. Biopolymer organization upon confinement. *J. Phys. Cond. Matt.*, 22:283102, 2010.

[87] J. F. Marko and E. D. Siggia. Stretching dna. *Macromolecules*, 28:8759–8770, 1995.

[88] W. F. Marshall, A. F. Dernburg, B. Harmon, D. A. Agard, and J. W. Sedat. Specific interactions of chromatin with the nuclear envelope: Positional determination within the nucleus in *Drosophila melanogaster*. *Molecular Biology of the Cell*, 7:825–842, 1996.

[89] G. G. Maul and L. Deaven. Quantitative determination of nuclear pore complexes in cycling cells with differing dna content. *The Journal of Cell Biology*, 73:748–760, 1977.

[90] R. Metzler, A. Hanke, P. G. Dommersnes, Y. Kantor, and M. Kardar. Tightness of slip-linked polymer chains. *Phys. Rev. Lett.*, 65:061103, 2002.

[91] R. Metzler, Y. Kantor, and M. Kardar. Force-extension relations for polymers with sliding links. *Phys Rev E*, 66:022102, 2002.

[92] D. Michieletto, M. Chiang, D. Coli, A. Papantonis, E. Orlandini, P. R. Cook, and D. Marenduzzo. Shaping Epigenetic Memory via Genomic Bookmarking. *Nucleic Acids Res.*, 46:83–93, 2018.

[93] D. Michieletto, E. Orlandini, and D. Marenduzzo. Polymer Model with Epigenetic Recolouring Reveals a Pathway for the de novo Establishment and 3D Organisation of Chromatin Domains. *Phys. Rev. X*, 6:041047, 2016.

[94] Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R. O'Meally, G. L. Dignon, A. E. Conicella, W. Zheng, R. B. Best, R. N. Cole, J. Mittal, F. Shewmaker, and N. L. Fawzi. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.*, 36:e201696394, 2017.

[95] F. Mueller, D. Mazza, T. J. Stasevich, and J. G. Mcnally. FRAP and kinetic modeling in the analysis of nuclear protein. *Curr. Opin. Cell Biol.*, 22:403–411, 2010.

[96] N. Naumova, M. Imakaev, G. Fudenberg, Y. Zhan, B. R. Lajoie, L. A. Mirny, and J. Dekker. Organization of the mitotic chromosome. *Science*, 342:948–953, 2013.

[97] M. Nicodemi and A. Pombo. Models of chromosome structure. *Current opinion in cell biology*, 28:90–95, 2014.

[98] M. Nicodemi and A. Prisco. Thermodynamic pathways to genome spatial organization in the cell nucleus. *Biophysical journal*, 96:2168–2177, 2009.

[99] E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, and B. G. Bruneau. Targeted degradation of ctcf decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169:930–944.e22, 2017.

[100] J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, and L. A. Mirny. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Biophysical Journal*, 114:30a, 2017.

[101] T. Odijk. Osmotic compaction of supercoiled dna into a bacterial nucleoid. *Biophysical Chemistry*, 73:23–29, 1998.

[102] A. Papantonis and P. R. Cook. Transcription factories: Genome organization and gene regulation. *Chemical Reviews*, 113:8683—-8705, 2013.

[103] L. A. Parada and T. Misteli. Chromosome positioning in the interphase nucleus. *Trends in Cell Biology*, 12:425–432, 2002.

[104] J. Pelletier, K. Halvorsen, B.-Y. Ha, Paparcone R., S. J. Sandler, C. L. Woldringh, W. P. Wong, and S. Jun. Physical manipulation of the escherichia coli chromosome reveals its soft nature. *Proceedings Natl. Acad. Sci. USA*, 109:E2649–E2656, 2012.

[105] M. C. F. Pereira, C. A. Brackley, J. S. Lintuvuori, D. Marenduzzo, and E. Orlandini. Entropic elasticity and dynamics of the bacterial chromosome: a simulation study. *J. Chem. Phys.*, 147:044908, 2017.

[106] R. Phillips, J. Kondev, J. Theriot, and H. Garcia. *Physical Biology of the Cell*. Garland Science, 2012.

[107] H. Pickersgill, B. Kalverda, E. de Wit, W. Talhout, M. Fornerod, and B. van Steensel. Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nature Genetics*, 38:1005–1014, 2006.

[108] S. A. Quinodoz, N. Ollikainen, B. Tabak, A. Palla, J. M. Schmidt, E. Detmar, M. M. Lai, A. A. Shishkin, P. Bhat, Y. Takei, V. Trinh, E. Aznauryan, P. Russell, C. Cheng, M. Jovanovic, A. Chow, L. Cai, P. Mcdonel, M. Garber, and M. Guttman. Higher-order inter-chromosomal hubs shape 3d genome organization in the nucleus. *Cell*, 174:744–757.e24, 2018.

[109] D. Racko, F. Benedetti, J. Dorier, and A. Stasiak. Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of tads in interphase chromosomes. *Nucleic Acids Research*, 46:1648–1660, 2017.

[110] S. Rao, S.-C. Huang, B. Glenn St. Hilaire, J. M. Engreitz, E. M. Perez, K.-R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, I. D. Bochkov, X. Huang, M. S. Shamim, A. D. Omer, B. E. Bernstein, R. Casellas, E. S. Lander, and E. Lieberman Aiden. Cohesin loss eliminates all loop domains, leading to links among superenhancers and downregulation of nearby genes. *Cell*, 171:305–320, 2017.

[111] S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. Lieberman Aiden. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159:1665–1680, 2014.

[112] C. H. Reccius, J. T. Mannion, J. D. Cross, and H. G. Craighead. Compression and free expansion of single dna molecules in nanochannels. *Phys. Rev. Lett.*, 95:268101, 2005.

[113] K. L. Reddy, J. M. Zullo, E. Bertolino, and H. Singh. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, 452:243–247, 2008.

[114] W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey, and R. H. Austin. Statics and dynamics of single dna molecules confined in nanochannels. *Phys. Rev. Lett.*, 94:196101, 2005.

[115] A. Rosa and R. Everaers. Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.*, 4:e1000153, 2008.

[116] A. Rosa, T. X. Hoang, D. Marenduzzo, and A. Maritan. Elasticity of semiflexible polymers with and without self-interactions. *Macromolecules*, 36:10095–10102, 2003.

[117] M. J. Rowley, M. H. Nichols, X. Lyu, M. Ando-Kuri, I. S. M. Rivera, K. Hermetz, P. Wang, Y. Ruan, and V. G. Corces. Evolutionarily conserved principles predict 3d chromatin organization. *Molecular Cell*, 67:837–852.e7, 2017.

[118] M. Rubinstein and R. H. Colby. *Polymer Physics*. Oxford University Press, 2003.

[119] K. Runcong and S. Mitaku. Comparative analysis of protein charge distribution in genome scale. *Genome Informatics*, 12:364–365, 2001.

[120] A. L. Sanborn, S. S. P. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Lia, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, and E. Lieberman Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences USA*, 112:E6456–E6465, 2015.

[121] S. Sati and G. Cavalli. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*, 126:33–44, 2017.

[122] W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Yann L.-M., N. A. Fonseca, W. Huber, C. Haering, L. Mirny, and F. Spitz. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551:51–56, 2017.

[123] V. F. Scolari and M. C. Lagomarsino. Combined collapse by bridging and self-adhesion in a prototypical polymer model inspired by the bacterial nucleoid. *Soft Matter*, 11:1677–1687, 2015.

[124] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148:458–472, 2012.

[125] T. N. Shendruk, M. Bertrand, H. W. de Haan, J. L. Harden, and G. W. Slater. Simulating the entropic collapse of coarse-grained chromosomes. *Biophys. J.*, 108:810–820, 2015.

[126] J. E. Sleeman and L. Trinkle-Mulcahy. Nuclear bodies: new insights into assembly/dynamics and disease relevance. *Current Opinion in Cell Biology*, 28:76–83, 2014.

[127] C. A. Sloan, E. T. Chan, J. M. Davidson, V. S. Malladi, J. S. Strattan, B. C. Hitz, I. Gabdank, A. K. Narayanan, M. Ho, B. T. Lee, L. D. Rowe, T. R. Dreszer, G. Roe, N. R. Podduturi, F. Tanaka, E. L. Hong, and J. M. Cherry. Encode data at the encode portal. *Nucleic Acids Research*, 44:D726–D732, 2016.

[128] D. Song and J. J. Loparo. Building bridges within the bacterial chromosome. *Trends Genet.*, 31:164–173, 2015.

[129] J. Stavans and A. Oppenheim. Dna-protein interactions and bacterial chromosome architecture. *Phys. Biol.*, 3:R1–R10, 2006.

[130] J. Stigler, G. Çamdere, D. E. Koshland, and E. C. Greene. Single-molecule imaging reveals a collapsed conformational state for dna-bound cohesin. *Cell Reports*, 15:988–998, 2016.

[131] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen. Phase separation drives heterochromatin domain formation. *Nature*, 547:241–245, 2017.

[132] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76:637, 1982.

[133] T. Terakawa, S. Bisht, J. M. Eeftens, C. Dekker, C. H. Haering, and E. C. Greene. The condensin complex is a mechanochemical motor that translocates along dna. *Science*, 358:672–676, 2017.

[134] S. S. Teves, L. An, A. S. Hansen, L. Xie, X. Darzacq, and R. Tjian. A dynamic mode of mitotic bookmarking by transcription factors. *Elife*, 5:1–24, 2016.

[135] N. M. Toan, D. Marenduzzo, P. R. Cook, and C. Micheletti. Depletion effects and loop formation in self-avoiding polymers. *Phys. Rev. Lett.*, 97:178302, 2006.

[136] N. L. van Berkum, E. Lieberman-Aiden, L. Williams, M. Imakaev, A. Gnirke, L. A. Mirny, J. Dekker, and E. S. Lander. Hi-c: A method to study the three-dimensional architecture of genomes. *J. Vis. Exp.*, 39:1869, 2010.

[137] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching dna with optical tweezers. *Biophysical Journal*, 72:1335–1346, 1997.

[138] A. H. Wani, A. N. Boettiger, P. Schorderet, A. Ergun, C. Munger, R. I. Sadreyev, X. Zhuang, R. E. Kingston, and N. J. Francis. Chromatin topology is coupled to polycomb group protein subnuclear organization. *Nature Communications*, 7:10291, 2016.

[139] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.*, 54:5237, 1971.

[140] A. S. Wegner, K. Wintraecken, R. Spurio, C. L. Woldringh, R. de Vries, and T. Odijk. Compaction of isolated *Escherichia coli* nucleoids: Polymer and h-ns protein synergetics. *J. Struct. Biol.*, 194:129–137, 2016.

[141] S. G. Wolf, D. Frenkiel, T. Arad, S. Finkel, R. Kolter, and A. Minsky. Dna protection by stress-induced biocrystallization. *Nature*, 400:83–85, 1999.

[142] M. Yu and B. Ren. The three-dimensional organization of mammalian genomes. *Annual Review of Cell and Developmental Biology*, 33:265–289, 2017.